

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Transcriptome Analysis of Non-Coding RNAs in Livestock Species: Elucidating the Ambiguity

---

Duy N. Do, Pier-Luc Dudemaine,  
Bridget Fomenky and Eveline M. Ibeagha-Awemu

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.69872>

---

## Abstract

The recent remarkable development of transcriptomics technologies, especially next generation sequencing technologies, allows deeper exploration of the hidden landscapes of complex traits and creates great opportunities to improve livestock productivity and welfare. Non-coding RNAs (ncRNAs), RNA molecules that are not translated into proteins, are key transcriptional regulators of health and production traits, thus, transcriptomics analyses of ncRNAs are important for a better understanding of the regulatory architecture of livestock phenotypes. In this chapter, we present an overview of common frameworks for generating and processing RNA sequence data to obtain ncRNA transcripts. Then, we review common approaches for analyzing ncRNA transcriptome data and present current state of the art methods for identification of ncRNAs and functional inference of identified ncRNAs, with emphasis on tools for livestock species. We also discuss future challenges and perspectives for ncRNA transcriptome data analysis in livestock species.

**Keywords:** bioinformatics, genome editing, livestock species, long non-coding RNA, non-coding RNA, microRNA, transcriptome

---

## 1. Introduction

A vast portion of the mammalian transcriptome is composed of non-protein coding transcripts or non-coding RNA (ncRNA). Some ncRNAs are processed into functionally important transcripts such as microRNA (miRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), small interfering RNA (siRNA), PIWI-interacting RNA (piRNA), circular RNA (circRNA), long non-coding RNA (lncRNA)

and several classes with limited information about their functions. In addition to the well described ncRNA classes, clusters of ncRNA (22–200 nucleotides (nt)) were detected at the 5' and 3' end of human and mouse genes, and named promoter-associated short RNAs (PASRs) and termini-associated short RNAs (TASRs) [1]. Mercer et al. [2] described a class of ncRNA, about 50–200 nt, that are processed from the 3'UTRs of protein-coding genes (uaRNAs). The uaRNAs are in sense direction to the protein-coding gene and show stage, sex and sub-cellular specific expression. A class of ncRNA derived from tRNA precursors and named tRNA-derived RNA fragments (tRF) or tRNA-derived small RNAs (tsRNAs) appear to be processed by Dicer while others are Dicer independently processed [3, 4]. Small nucleolar RNAs (snoRNA) can also be processed into small miRNA-like molecules called sno-derived RNAs or sdRNAs [5, 6] which play roles in guiding enzymes to target RNAs for modification [7]. In this chapter, only the main classes of functional ncRNAs (miRNA, snoRNA, siRNA, piRNA and lncRNA), not considering the translation related ncRNAs (rRNA and tRNA), will be further discussed. NcRNAs have been implicated in many biological processes including transcriptional inference, translational modifications, mRNA cleavage, epigenetic modifications, regulation of structural organization, and modulation of alternative splicing, small RNA precursor, and endo or secondary siRNA generation [7–10].

## 2. Transcriptome analysis of non-coding RNA

### 2.1. Platforms for transcriptome analysis of non-coding RNA

Transcriptome analysis reached a turning point in its history with the arrival of high throughput next-generation sequencing technologies like RNA-Sequencing (RNA-Seq) [11, 12]. Before this time, microarray was the gold standard for transcript profiling or simultaneous measurement of the expression level of thousands of genes in a given sample [13, 14]. Microarray technology however has major drawbacks like non-specific probe hybridization signals and errors in background level measurements [15], as well as limited gene diversity since probes are designed to represent only a set of preselected genes. Unique hybridization properties of each probe may affect their dynamic range and thus create bias in data processing algorithms [16]. The flexibility offered by RNA-Seq technology enables detection of unknown splice junctions [17], novel transcripts [18], new single nucleotide polymorphisms (SNPs) [19] and many other features all in the same assay. RNA-Seq technology has taken the possibility of fine tuning our knowledge of the transcriptome to a much higher level. In recent years, RNA-Seq has proved its worth as a technology that will replace microarray in whole-genome transcript profiling [20–22]. Correlation of RNA-Seq to RNA-Seq differential gene expression data resulted in good overlap than RNA-Seq to microarray data [23, 24], thus confirming that RNA-Seq is the preferred method to analyze the transcriptome. Moreover, correlation of transcriptome quantification by the two methods versus transcript level measured by shotgun mass spectroscopy showed better estimation with RNA-Seq analysis [25]. Through the evolution process of RNA-Seq technology, other new aspects have been included such as allele specific transcriptome analysis. Moreover, since the RNA-Seq procedure does not rely

on known genome annotation, but rather on all the information available in a given sample, there is clear opportunity to make discoveries at a rate never expected before.

A diversity of platforms offer a wide range of RNA-sequencing possibilities[12]. For example, Illumina HiSeq and MiSeq technologies offer short sequence reads (36–300 base pairs (bp)) while Oxford Nanopore can reach sequence lengths of greater than 150 kilo base pairs (kb) [26]. The sequencing techniques could be DNA-polymerase dependent (i.e. sequencing-by-synthesis (e.g. Illumina MiSeq/HiSeq)) while others like PacBio and Oxford Nanopore are single-molecule sequencers. The sequencing error rate ranges from 0.1% (Illumina MiSeq/HiSeq) to about 1.3% (PacBio RSII single pass). An overview of sequencing platforms and their characteristics is shown in **Table 1**. The error rate between platforms varies [27], so it is important to consider this especially when the goal is to sequence short read transcripts like miRNA.

The challenges of managing RNA-Seq data are considerable in terms of data storage and analysis as well as algorithm development. Since the technology is not yet fully matured, shortcomings exist at every step of sequence analysis. Various tools are available for alignment of reads, transcript construction, quantification, differential gene expression, pathways and correlation analyses [28] (**Tables 2 and 3**). Nonetheless, the use and specificity of the softwares differ highly from one type of analysis to another and the hardest part is making sure that the right tool is chosen at every step. A review of best practices for RNA-Seq data analysis was published recently [29]. The gap between the rapid evolution of RNA-Seq technology and the development of data analysis tools is hindering wide application in livestock species. Most data analysis tools are developed for use with genomes of human and common model organisms (mouse, rat) and require tweaking before use with livestock genomes. For example, when performing target prediction analysis for newly discovered transcripts, it is the practise to use human/mouse databases as it brings a lot of power to the analysis. However, there is great bias coming from the assumption that livestock biological systems are identical to human or mouse.

## 2.2. Generation of ncRNA sequence data and pre-mapping quality control

### 2.2.1. Generation of ncRNA sequence data

The choice of the sequencing platform is critical to attain the goals of a study. Numerous protocols and commercial kits to generate cDNA libraries from RNA samples are available and they are mostly based on the same principles (e.g. fragmentation, reverse-transcription, adapter ligation and amplification). The steps in library preparation for lncRNA are the same as for mRNA since they share similar biogenesis pathways. The starting material for lncRNA library preparation is total RNA. Majority of lncRNA transcripts have poly-A tails while a small proportion do not. Library preparation methods based on poly-A tail selection are cheaper but less robust since non-poly-A tail transcripts are lost. An ideal but more expensive method involves depletion of rRNA (constitutes ~90% of total RNA). Library preparation with rRNA depleted total RNA is robust as it allows quantification of all other RNA transcripts including lowly expressed transcripts. Thus, the first step in lncRNA library preparation is to consider whether to perform poly-A tail selection or to deplete rRNA (**Figure 1**). The next dilemma is deciding whether or not

Platform	Read length <sup>1</sup> (base pair)	Throughput <sup>2</sup>	Number of reads <sup>3</sup>	Error profile
Illumina MiniSeq (high output)	75 (SE)	1.6–1.8 Gb	22–25 M	<1%, substitution
	75 (PE)	3.3–7.5 Gb	44–50 M	<1%, substitution
	150 (PE)	6.6–7.5 Gb	44–50 M	
Illumina MiniSeq (mid output)	75 (SE)	2.1–2.4 Gb	14–16 M	<1%, substitution
Illumina MiSeq v2	36 (SE)	540–610 Mb	12–15 M	<0.1%, substitution
	25 (PE)	750–850 Mb	24–30 M	<0.1%, substitution
	150 (PE)	4.5–5.1 Gb	24–30 M	<0.1%, substitution
	250 (PE)	7.5–8.5 Gb	24–30 M	<0.1%, substitution
Illumina MiSeq v3	75 (PE)	3–4 Gb	44–50 M	<0.1%, substitution
	300 (PE)	13–15 Gb	44–50 M	<0.1%, substitution
Illumina NextSeq 500/550 (high output)	75 (SE)	25–30 Gb	400 M	<1%, substitution
	75 (PE)	50–60 Gb	800 M	<1%, substitution
	150 (PE)	100–120 Gb	800 M	<1%, substitution
Illumina NextSeq 500/550 (mid output)	75 (PE)	16–20 Gb	~260 M	<1%, substitution
	150 (PE)	32–40 Gb	~260 M	<1%, substitution
Illumina HiSeq250v2 Rapid run	36 (SE)	9–11 Gb	300 M	0.1%, substitution
	50 (PE)	25–30 Gb	600 M	0.1%, substitution
	100 (PE)	50–60 Gb		0.1%, substitution
	150 (PE)	75–90 Gb		0.1%, substitution
	250 (PE)	125–150 Gb		0.1%, substitution
Illumina HiSeq250v3	36 (SE)	47–52 Gb	1.5 B	0.1%, substitution
	50 (PE)	135–150 Gb	3 B	0.1%, substitution
	100 (PE)	270–300 Gb		0.1%, substitution
Illumina HiSeq250v4	36 (SE)	64–72 Gb	2 B	0.1%, substitution
	50 (PE)	180–200 Gb	4 B	0.1%, substitution
	100 (PE)	360–400 Gb		0.1%, substitution
	125 (PE)	450–500 Gb		0.1%, substitution
Illumina HiSeq3000/4000	50 (SE)	105–125 Gb	2.5 B	0.1%, substitution
	75 (PE)	325–375 Gb		0.1%, substitution
	150 (PE)	650–750 Gb		0.1%, substitution
Illumina HiSeqX	150 (PE)	800–900 Gb	2.6–3 B	0.1%, substitution
	150 (PE)	1.6–20 B	167 Gb–6 Tb	

Platform	Read length <sup>1</sup> (base pair)	Throughput <sup>2</sup>	Number of reads <sup>3</sup>	Error profile
Ion Proton	200 (SE)	Up to 10 Gb	60 M	1% indel
Ion PGM 318	200 or 400 (SE)	0.6–2 Gb	4–5.5 M	1% indel
Ion PGM 316	200 or 400 (SE)	0.3–1 Gb	2–3 M	1% indel
Ion PGM 314	200 or 400 (SE)	30–100 Mb	0.4–0.5 M	1% indel
PacBio Sequel	8–12 kb (SE)	3.5–7 Gb	>100,000	N/A
PacBio RS II	~20 kb	0.5–1Gb	~55,000	~13%, indel
454 GS Junior	~400 (SE, PE)	35 Mb	~0.1 M	1%, indel
454 GS Junior+	~700 (SE, PE)	70 Mb	~0.1 M	1%, indel
454 GS FLX Titanium XLR70	Up to 600; 450 mode (SE, PE)	450 Mb	~1 M	1%, indel
454 GS FLX Titanium XL+	Up to 1000; 700 mode (SE, PE)	700 Mb	~1 M	1%, indel
SOLiD 5500 xl	50 or 75 (SE)	160–320 Gb	~1.4 B	≤0.1%, AT bias
SOLiD 5500 Wildfire	50 or 75 (SE)	80–160 Gb	700 M	≤0.1%, AT bias
Oxford Nanopore MK1 MinION	Up to 200 Kb	~1.5 Gb		~12%, indel
Oxford Nanopore GridION X5	~Hundreds of Kb	100 Gb		
Oxford Nanopore PromethION		~4 Tb		

<sup>1</sup>SE: single end, PE: paired end, Kb, Kilo base pair.

<sup>2</sup>Mb: Megabyte, Gb: Gigabyte, TB: Terabyte.

<sup>3</sup>M: Million, B: Billion.

**Table 1.** Overview of some sequencing platforms for transcriptome analysis and their characteristics.

to preserve strand information during library preparation. As lncRNA annotation is still in the initial phase, it is crucial to preserve strand information to enable correct genome localization of novel transcripts. Paired-end sequencing is to be considered over single end sequencing for lncRNA characterization to facilitate construction of transcripts with clear-cut exon boundaries. Paired-end sequencing also allows accurate detection of splicing position. Sequencing long fragments (>100 bp) is also desired to get adequate coverage of the genome and consequently, better transcript construction. The number of multiplexed samples on each sequencing lane affects lncRNA sequence depth. Reducing cost by multiplexing more samples than necessary reduces quality of results obtained. It has been demonstrated that the depth of sequencing is relative to the nature of the expected results [30, 31]. To accomplish lncRNA discovery with confidence, a minimum of 100 million reads per sample is suggested to enable *de novo* transcript assembly.

The procedure for the generation of miRNA sequence data differs slightly from the procedure for lncRNA analysis. First of all, miRNAs are small (18–24 bp) in size and do not require RNA



Step	Tools	Application/Web link	References
Trimming*	Trimmomatic	Illumina single end and paired end quality and adapter trimming. <a href="http://www.usadellab.org/cms/?page=trimmomatic">http://www.usadellab.org/cms/?page=trimmomatic</a>	[39]
	PEAT	Specific for paired end sequencing quality and adapter trimming. <a href="https://github.com/jhhung/PEAT">https://github.com/jhhung/PEAT</a>	[50]
	Trim Galore	Quality and adapter trimming with some extra functionality for Bisulfite-Seq. <a href="https://www.bioinformatics.babraham.ac.uk/projects/trim_galore">https://www.bioinformatics.babraham.ac.uk/projects/trim_galore</a>	[51]
	Skewer	Adapter trimming, can take into account indels. <a href="https://github.com/relipmoc/skewer">https://github.com/relipmoc/skewer</a>	[52]
	AlienTrimmer	Detect and remove alien k-mers in both ends of sequence reads. <a href="ftp://ftp.pasteur.fr/pub/gensoft/projects/AlienTrimmer/">ftp://ftp.pasteur.fr/pub/gensoft/projects/AlienTrimmer/</a> .	[53]
	Cutadapt	Finds and remove adapter, primers, poly-A and other types of unwanted sequences. <a href="https://github.com/marcelm/cutadapt">https://github.com/marcelm/cutadapt</a>	
	NxTrim	Discard as little sequence as possible from Illumina Nextera Mate Pair reads, single end and paired end reads. <a href="https://github.com/sequencing/NxTrim">https://github.com/sequencing/NxTrim</a>	[54]
	SeqPurge	Can detect very short adapter sequences. <a href="https://github.com/imgag/ngs-bits/blob/master/doc/tools/SeqPurge.md">https://github.com/imgag/ngs-bits/blob/master/doc/tools/SeqPurge.md</a>	[55]
Alignment**	STAR	Align RNA-Seq reads to a reference genome, detect splice junctions. <a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>	[45]
	Bowtie / Bowtie2	Align short DNA sequences to genomes with Burrows-Wheeler index. <a href="http://bowtie-bio.sourceforge.net/bowtie2">bowtie-bio.sourceforge.net/bowtie2</a>	[56, 57]
	BWA	Mapping low-divergent sequences against large reference genome. <a href="http://bio-bwa.sourceforge.net">bio-bwa.sourceforge.net</a>	[58]
	TopHat2	Use Bowtie for alignment. TopHat analyzes results to identify splice junctions. <a href="https://ccb.jhu.edu/software/tophat">https://ccb.jhu.edu/software/tophat</a>	[59]
	Rockhopper	Specific for bacterial RNA-Seq data. It supports de novo and reference based transcript assembly. <a href="http://cs.wellesley.edu/~btjaden/Rockhopper">cs.wellesley.edu/~btjaden/Rockhopper</a>	[60]
	SpliceMap	De novo splice junction discovery and alignment tool. <a href="https://web.stanford.edu/group/wonglab/SpliceMap">https://web.stanford.edu/group/wonglab/SpliceMap</a>	[61]
	StringTie	De novo transcript assembly. Quantitation of full-length transcripts representing multiple splice variants for each gene locus. <a href="https://ccb.jhu.edu/software/stringtie">https://ccb.jhu.edu/software/stringtie</a>	[47]
	Trinity	De novo reconstruction of transcriptomes from RNA-seq data. <a href="https://github.com/trinityrnaseq/trinityrnaseq/wiki">https://github.com/trinityrnaseq/trinityrnaseq/wiki</a>	[62]

\*Further trimming tools are available at: <https://omictools.com/adapter-trimming-category/>

\*\*Further alignment tools are available at: <https://omictools.com/read-alignment-category/>

Table 2. Frequently used tools for trimming and alignment.

Names	Major purpose <sup>1</sup>	Known miRNA annotation <sup>2</sup>	Novel miRNA discovery	DE analyses	Target prediction	Pathway enrichment	Livestock Species	References
miRDeep	miRNA identification	+	+	–	–	–	+	[74]
mirTools	miRNA identification	+	+	+	+	–	+	[71]
UEA sRNA Workbench	miRNA identification	+	+	+	+	–	+	[76]
sRNAtoolbox	miRNA identification	+	+	+	+	–	+	[77]
MiReNA	miRNA identification	+	–	–	–	–	+	[81]
miRExpress	miRNA identification	–	+	–	–	–	+	[93]
DARIO	miRNA identification	+	+	–	+	–	–	[94]
Target scan	Target prediction	–	–	–	+	–	+	[95]
DIANA-microT- CDS	Target prediction	–	–	–	+	+	–	[96]
miRanda	Target prediction	–	–	–	+	–	+	[97]
miRDB	Target prediction	–	–	–	+	–	+	[98]
miRTar	Target prediction	–	–	–	+	–	–	[99]
mirWIP	Target prediction	–	–	–	+	–	–	[100]
MMIA	Target prediction	–	–	–	+	+	–	[101]
PITA	Target prediction	–	–	–	+	–	+	[102]
psRNATarget	Target prediction	–	–	–	+	–	–	[103]
RNA22	Target prediction	–	–	–	+	–	+	[104]
RNAhybrid	Target prediction	–	–	–	+	–	+	[105]



Names	Major purpose <sup>1</sup>	Known miRNA annotation <sup>2</sup>	Novel miRNA discovery	DE analyses	Target prediction	Pathway enrichment	Livestock Species	References
TargetRank	Target prediction	–	–	–	+	–	–	[106]
DIANA-mirPath v3	Down-stream miRNA analyses	–	–	–	+	+	+	[107]
miRGator	Integrated tools	–	–	+	+	+	–	[108]
MAGIA	Down-stream miRNA analyses	–	–	–	–	+	–	[109]
miRNet	Down-stream miRNA analyses	–	–	–	–	+	+	[110]
miRSystem	Down-stream miRNA analyses	–	–	–	–	+	+	[111]
miRNAMap	Integrated tools	+	+	–	+	+	+	[112]
miRTarBase	Integrated tools	+	+	+	+	+	+	[113]
TransmiR	Down-stream miRNA analyses	–	–	–	–	+	+	[114]
PicTar	Target prediction	–	–	–	+	–	+	[115]
miRWalk	Integrated tools	+	+	+	–	–	+	[116]
MiRecords	Integrated tools	+	+	+	–	–	+	[117]
multiMiR	Integrated tools	+	+	+	–	–	+	[118]
miRconnX	Integrated tools	+	+	+	+	+	–	[119]
DIANA-mirExTra	Down-stream miRNA analyses	–	–	–	–	+	–	[120]
TarBase	Database	+	+	+	–	+	+	[121]

<sup>1</sup>Further tools for miRNA annotation are available at: [https://tools4mirs.org/software/known\\_mirna\\_identification/](https://tools4mirs.org/software/known_mirna_identification/); Further tools for novel miRNA discovery and miRNA precursor prediction are available at: [https://tools4mirs.org/software/precursor\\_prediction/](https://tools4mirs.org/software/precursor_prediction/); Further tools for miRNA target prediction are available at: [https://tools4mirs.org/software/target\\_prediction/](https://tools4mirs.org/software/target_prediction/); <https://omictools.com/mirna-target-prediction-category>

<sup>2</sup>“+” Function is included, “–” Function is not included.

**Table 3.** Overview of tools used for the analysis of miRNA sequence data.

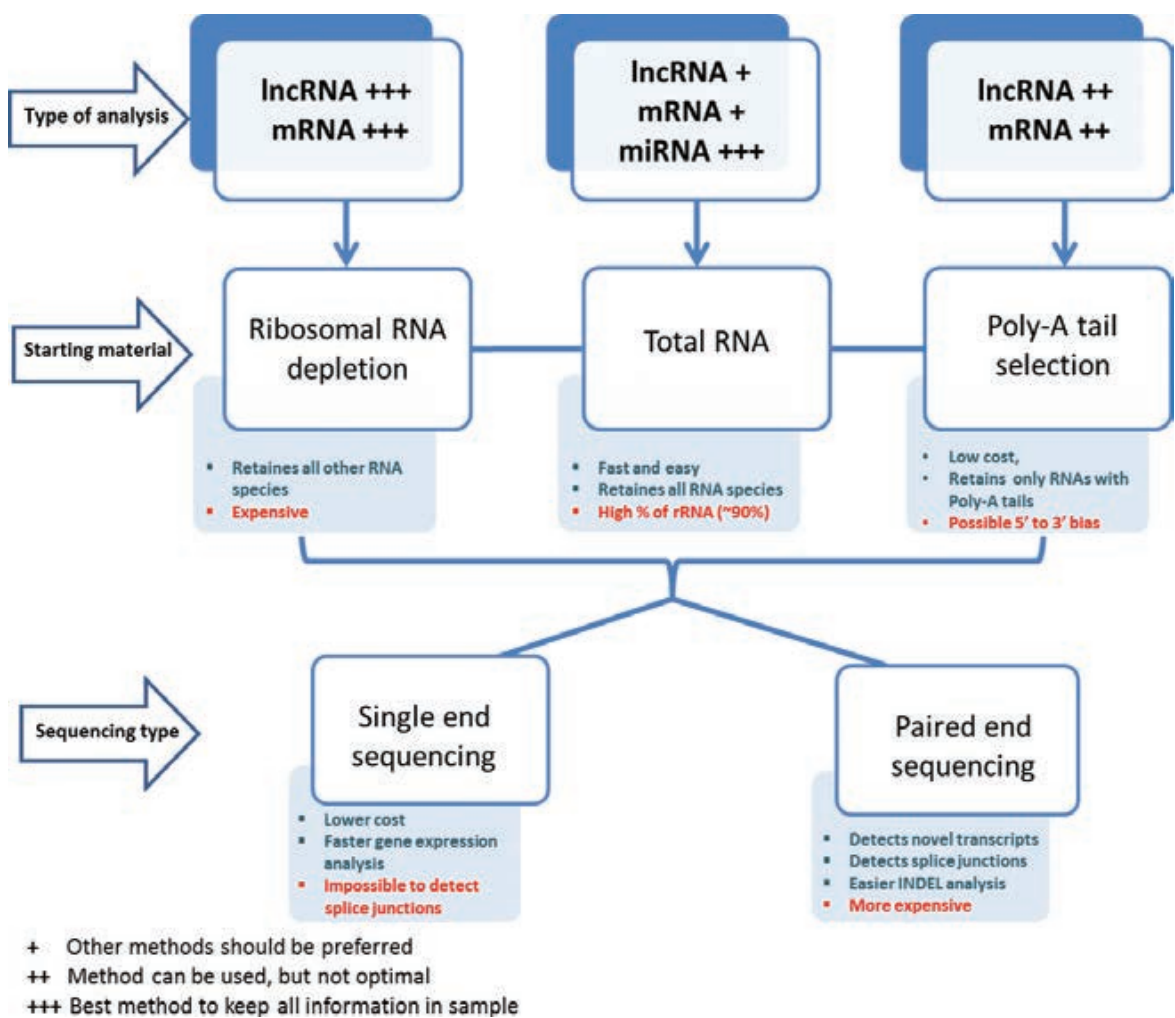


Figure 1. Starting material and sequencing method considerations according to RNA species to be analyzed.

fragmentation prior to library construction. Total RNA is the recommended starting material for miRNA library preparation (Figure 1). Although some commercial kits provide the option to enrich the miRNA fraction prior to library preparation, there is evidence that some small RNA species are lost during enrichment [32]. The protocols for miRNA library preparation are generally similar to lncRNA and include adapter ligation step, reverse transcription and amplification followed by size selection and purification of the cDNA. Fifty bp single end sequencing is sufficient for miRNA libraries since miRNAs are generally small. Thus, Illumina platforms are well suited for sequencing miRNA libraries. Studies showed that approximately 2 million reads are sufficient for differential expression analysis while 8 million reads are sufficient for discovery analysis [33, 34]. Considering that over 150 million reads are available per lane on HiSeq machines, sample multiplexing can be as high as 18 to 20 libraries per lane.

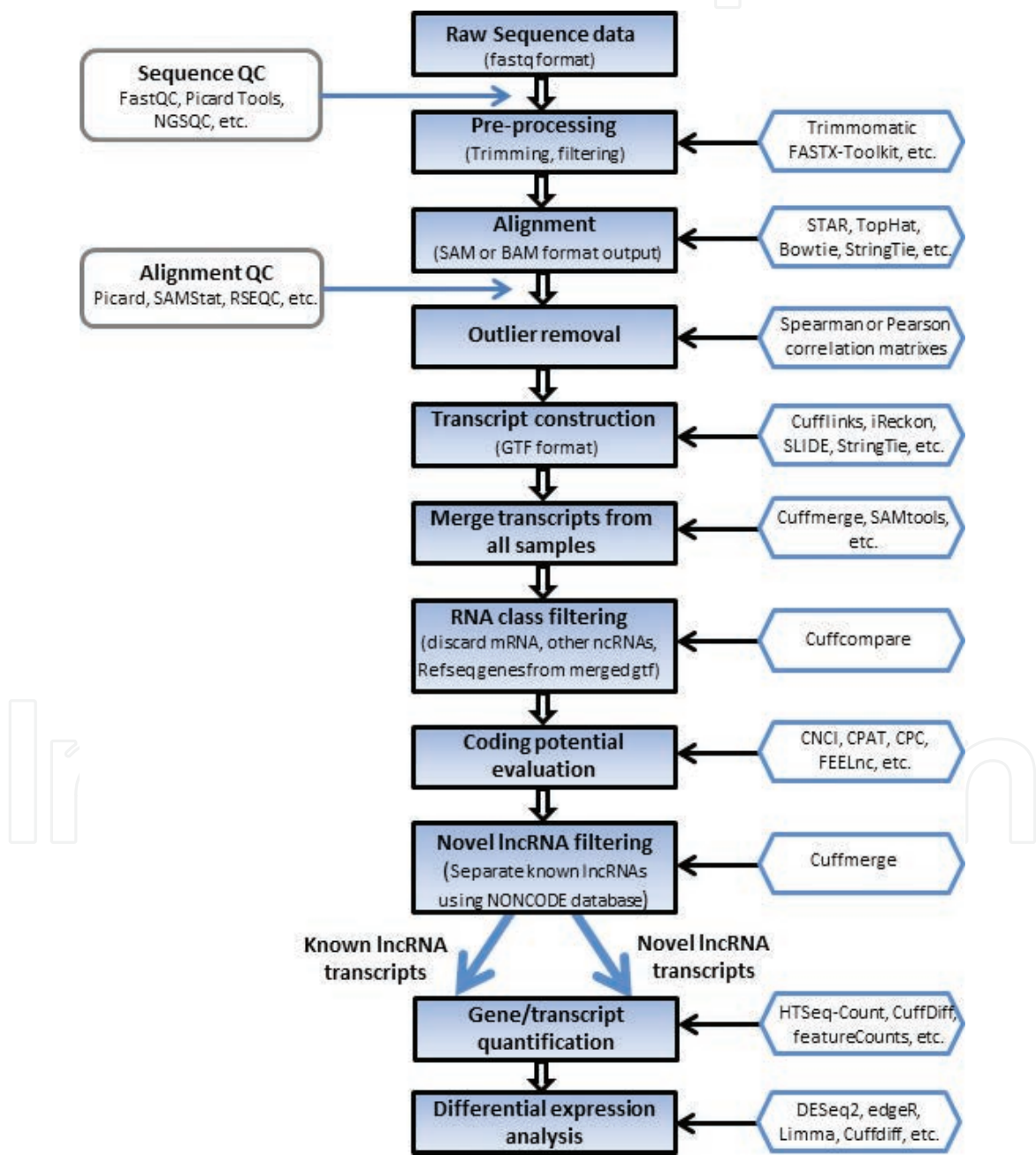
2.2.2. Common data processing steps

Upon availability of sequence data, many bioinformatics tools are used in the analytical procedures. Some processing steps are optional but strongly recommended; while others are required before the next step can be performed. Many pipelines have been developed to

answer specific questions, but the softwares used can be very different. A global view of the general processing steps and frequently used tools for lncRNA and miRNA sequence data analyses are presented in **Figures 2** and **3**, respectively. These processing steps can be modified to include desired or specific tools depending on the research question.

2.2.3. Raw data quality control

Sequence data generated by Illumina platforms and most platforms is in FASTQ format. The FASTQ format is a text file consisting of the nucleic acid sequence (read) and base calling accuracy score (Phred score) attributed to each base pair of the sequence. FastQC [35], Picard



**Figure 2.** General processing steps and tools used in lncRNA sequence analysis.

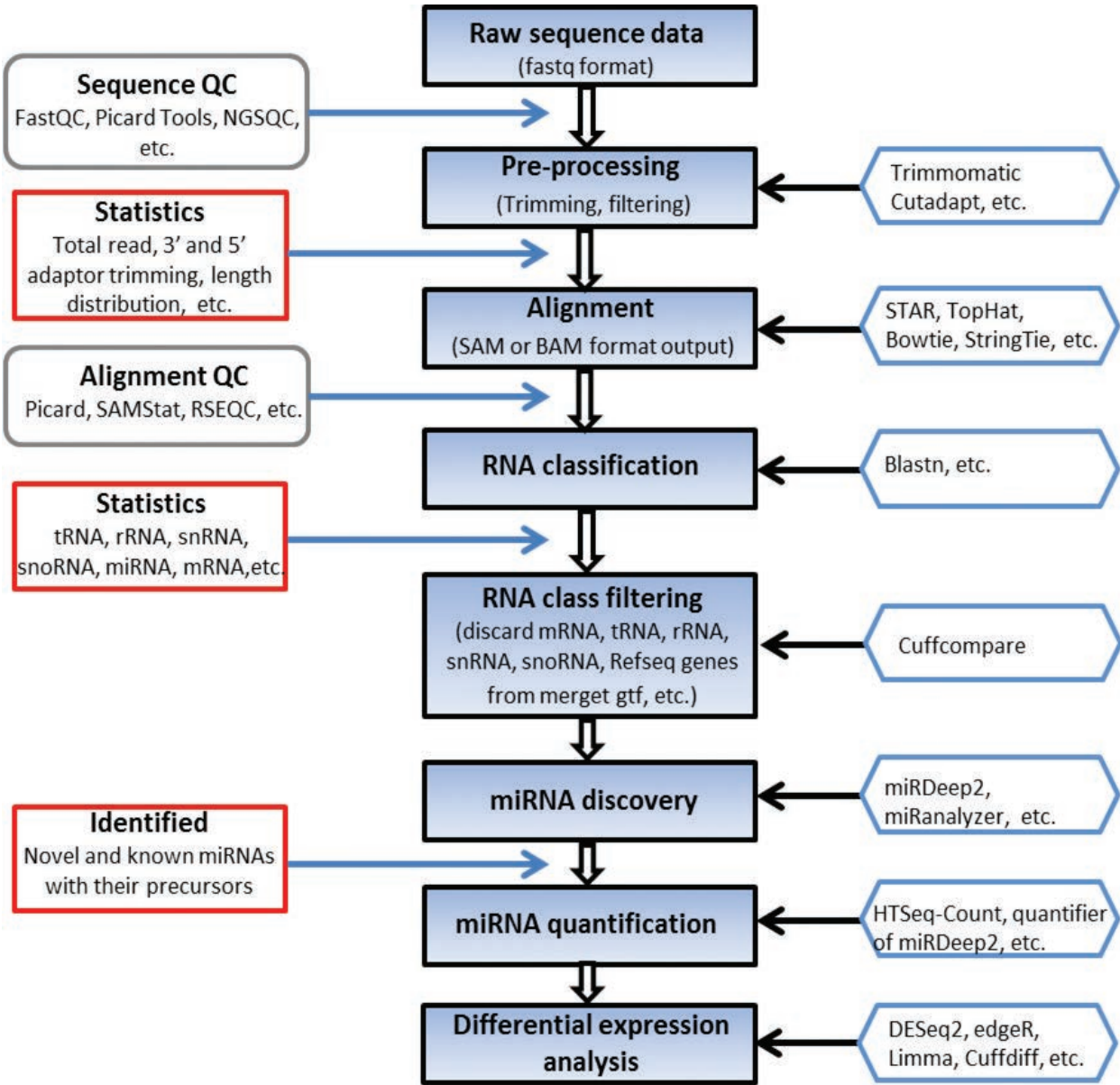


Figure 3. General processing steps and tools used in miRNA sequence analysis.

tools (<https://broadinstitute.github.io/picard/>) and NGS QC tool kit [36] are often used to assess the quality of raw sequence reads. This step is necessary to determine if the sequencing outcome is as expected. These tools inform on the total number of reads, the overall quality of base call according to the position, GC percentage and other features. Care should be taken when interpreting the results because GC content is species specific and some softwares evaluate GC content according to the human genome. In order to avoid bias in the mapping step, a quality trimming is necessary to get rid of low quality base pairs and remaining adapter sequences. A recent study showed that incorrect trimming can lead to generation of short reads impairing the capacity to correctly predict differences in expression changes [37]. Several trimming tools are available [38] (<https://omictools.com/adaptor-trimming-category>) including Trimmomatic [39], FASTX-Toolkit [40], CutAdapt [41], etc. (Table 2). Following trimming, filtering of reads is necessary to get rid of very short and overall low quality reads to keep bias level as low as possible.



#### 2.2.4. Alignment

After trimming and filtering, reads are ready for alignment or *de novo* construction. Alignment consists of mapping reads to a reference genome. Various alignment tools have been developed [42, 43] (<https://omictools.com/read-alignment-category>) including frequently used tools like TopHat [44], STAR [45], Bowtie [46], StringTie [47], etc. (**Table 2**). These softwares have their own specifications highlighting the importance of understanding the utility of each tool and the options they offer. The alignment tool used can have great impact on the end results. It has been observed that the choice of aligner and specific options can affect results of differential gene expression analysis [48]. Aligners can be grouped in two types, gapped (also known as split, e.g. STAR, BWA, etc.) and ungapped (e.g. Bowtie, etc.). Bowtie (ungapped group) can easily map reads to a genome, but is less effective at finding spliced junctions. Aligners in the gapped group are able to align reads and detect spliced variants. In the absence of a reference genome, *de novo* assembly aligners (e.g. Trinity [49]) can be used. In the context of lncRNA read alignment, gapped softwares are preferred since the transcripts are not all annotated and portions of the reads of the same transcript may align to one position of the genome and the remaining to another position. Alignment is one of the longest steps in RNA-Seq sequence analysis therefore selection of the right tool might have significant impact on the outcome of the analysis. It is also important to perform mapping quality control following alignment. Quality check includes the percentages of mapped and unmapped reads, the location of the reads (intronic and exonic) and the 5'–3' coverage.

#### 2.2.5. Transcript construction and quantification

RNA-Seq transcript construction and the alignment steps can demand considerable computing time. Transcript construction tools are many (<https://omictools.com/transcript-quantification-category>) including commonly used tools like Cufflinks [63], iReckon [64], StringTie [47], etc. This step requires paired-end data and high sequence coverage to reconstruct lowly expressed transcripts. With the assumption that transcripts are species specific, raw data or alignment files from all samples from the same population can be merged to increase coverage [65]. This modification will help clarify transcript boundaries in case of *de novo* transcript assembly. Particular considerations for lncRNA transcript construction include sample pooling according to species and tissue type. LncRNA expression is known to demonstrate tissue specificity [66–68].

#### 2.2.6. miRNA processing steps

Overall, the procedures for miRNA identification and discovery are less time consuming and do not include as many steps as for mRNA and lncRNA identification. The global process includes quality and adaptors trimming with quality checkpoints before and after each step. A size selection to keep sequences between 17 and 30 nt (sometimes up to 35 nt) is often performed right after the quality and adaptors trimming step. This is followed by read mapping and filtering of other RNA sequences (rRNA, tRNA, snRNA, mRNA, lncRNA, etc.). The reads thought to represent miRNA are analyzed with miRNA prediction tools like miRDeep2 [69],

miRanalyzer [70], mirTools 2.0 [71], etc. (**Table 3**). Subsequent interrogation of miRBase database enables classification of retained miRNAs as known or novel miRNAs. A tool like miRDeep2 has a quantifier module that generates a read count table for each miRNA using precursor and mature sequence files as input. An overview of tools for miRNA identification are presented in **Table 3** and further discussed in the next section.

### 3. Tools for ncRNA identification

#### 3.1. Tools for miRNA identification

The identification of miRNAs can be either annotation of known miRNAs or discovery of novel miRNAs. A variety of algorithms and bioinformatics tools are applied to annotate known miRNAs as well as to discover new miRNAs from sequence data. These tools can use several features such as sequence conservation among species, structural features like hairpin and minimal folding free energy [72]. Many tools are available for miRNA annotation ([https://tools4mirs.org/software/known\\_mirna\\_identification/](https://tools4mirs.org/software/known_mirna_identification/)) [73] including frequently used tools like miRdeep [74], miRanalyzer [75], mirTools 2.0 [71], UEA sRNA Workbench [76], sRNAtoolbox [77], and SeqBuster [78] (**Table 3**). Many more tools have been developed for novel miRNA discovery and miRNA precursor prediction ([https://tools4mirs.org/software/precursor\\_prediction/](https://tools4mirs.org/software/precursor_prediction/)) [73] including frequently used tools like MiPred [79], miRanalyzer [75], miR-Abela [80], MiReNA [81], UEA sRNA Workbench [76] and mirDeep [74] (**Table 3**). Major features of miRNA discovery tools have been reviewed [82–84]. Regarding livestock species, the choice of methods for miRNA discovery and novel miRNA annotation vary among studies and species. For example, De Vlieghe et al. [85] used miRbase [86] and UNAFold [87] for miRNA annotation and discovery in bovine mammary gland tissues while Peng et al [88] used miRbase [86] and RNAfold [89] for these purposes in porcine mammary glands. In our own studies, miRbase [86] and mirDeep2 [74] were used to identify miRNAs in various tissues including bovine mammary gland tissues [90], milk fat [90–92], milk whey and cells [90].

#### 3.2. Tools for lncRNA identification

To date, a large number of lncRNA genes have been identified in the genomes of human (141,353), cow (23,896) and chicken (13,085) (<http://www.bioinfo.org/noncode/analysis.php>, accessed on 24-03-2017). Several methodologies have been described to identify/distinguish lncRNAs from mRNAs and successfully applied to livestock species such as coding potential calculator (CPC) [122], PhyLoCSF [123], coding-non-coding index (CNCI) [124], coding potential assessment tool (CPAT) [125], Predictor of Long non-coding RNAs and mRNAs based on an improved k-mer scheme (PLEK) [126] and Flexible Extraction of LncRNAs (FEELnc) [127], etc. The FEELnc program developed by the functional annotation of animal genome project consortium (FAANG) [128] is recommended as a standardized protocol for lncRNA analyses in animal species. In order to distinguish lncRNAs from mRNAs, FEELnc program uses a machine-learning method for estimation of a protein-coding score according to the



RNA size, open reading frame coverage and multi k-mer usage [127]. The FEELnc program can derive an automatically computed cut-off so it maximizes the lncRNA prediction sensitivity and specificity. An overview of tools for lncRNA identification/characterization is listed in **Table 4**.

Tools	Type	Major Function/web link	References
ChIPBase	Database	Identifies binding motif matrices and their binding sites. Predicts transcriptional regulatory relationships between transcription factors and genes. <a href="http://rna.sysu.edu.cn/chipbase/">http://rna.sysu.edu.cn/chipbase/</a> .	[129]
LNCipedia	Database	Provides basic transcript information and structure, human lncRNA transcripts and genes. <a href="http://www.lncipedia.org/">http://www.lncipedia.org/</a> .	[130]
lncRNADB	Database	Provides comprehensive annotation of eukaryotic lncRNAs. Offers an improved user interface enabling greater accessibility to sequence information, expression data and the literature. <a href="http://www.lncrnadb.org/">http://www.lncrnadb.org/</a> .	[131]
LNCat	Database	Stores the information of 24 lncRNA annotation resources. Allows achieving refined annotation of lncRNAs within the interested region. <a href="http://biocc.hrbmu.edu.cn/LNCat/">http://biocc.hrbmu.edu.cn/LNCat/</a>	[132]
lncRNASNP	Database	Provide comprehensive resources of single nucleotide polymorphisms (SNPs) in human/mouse lncRNAs. <a href="http://bioinfo.life.hust.edu.cn/lncRNASNP/">bioinfo.life.hust.edu.cn/lncRNASNP/</a>	[133]
lncRNAWiki	Database	Provide open-content and publicly editable curation and collection of information on human lncRNAs. <a href="http://lncrna.big.ac.cn/index.php/Main_Page">http://lncrna.big.ac.cn/index.php/Main_Page</a>	[134]
NONCODE	Database	Presents the most complete collection and annotation of non-coding RNAs (excluding tRNAs and rRNAs) for 18 species including human, mouse, cow, rat, chicken, pig, fruitfly, zebrafish, <i>Caenorhabditis elegans</i> and yeast. <a href="http://www.noncode.org/">www.noncode.org/</a>	[135]
ALDB	Database	Enables the exploration and comparative analysis of lncRNAs in domestic animals. Offers information on genome-wide expression profiles and animal quantitative trait loci (QTLs) of domestic animals. <a href="http://res.xaut.edu.cn/aldb/index.jsp">http://res.xaut.edu.cn/aldb/index.jsp</a>	[136]
GENCODE	Database	Presents all gene features in the human genome. Contains annotation of lncRNA loci publicly available with the predominant transcript form consisting of two exons. <a href="https://www.gencodegenes.org">https://www.gencodegenes.org</a>	[137]

Tools	Type	Major Function/web link	References
ncRDeathDB	Database	Present a comprehensive bioinformatics resource to ncRNA-associated cell death interactions. <a href="http://www.rna-society.org/ncrdeathdb">www.rna-society.org/ncrdeathdb</a>	[138]
LncVar	Database	Presents genetic variation associated with long noncoding genes. <a href="http://bioinfo.ibp.ac.cn/LncVar">bioinfo.ibp.ac.cn/LncVar</a>	[139]
IRNdb	Database	Combines microRNA, PIWI-interacting RNA, and lncRNA information with immunologically relevant target genes. <a href="http://irnodb.org">http://irnodb.org</a>	[140]
AnnoLnc	Annotation	Presents online portal for systematically annotating newly identified human lncRNAs.	[141]
LongTarget	Target prediction	Present a computational method and program to predict lncRNA DNA-binding motifs and binding sites. <a href="http://lncrna.smu.edu.cn">lncrna.smu.edu.cn</a>	[142]
LncRNA2Function	Functional inferences	Facilitates search for the functions of a specific lncRNA or the lncRNAs associated with a given functional term, or annotate functionally a set of human lncRNAs of interest. <a href="http://mlg.hit.edu.cn/lncrna2function">http://mlg.hit.edu.cn/lncrna2function</a>	[143]
Co-LncRNA	Function inference	Presents a web-based computational tool that allows users to identify GO annotations and KEGG pathways that may be affected by co-expressed protein-coding genes of single or multiple lncRNAs. <a href="http://www.bio-bigdata.com/Co-LncRNA/">www.bio-bigdata.com/Co-LncRNA/</a>	[144]
LncReg	Function inference	Provides regulatory information about lncRNAs, such as targets, regulatory mechanisms, and experimental evidence for regulation and key molecules participating in regulation. <a href="http://bioinformatics.ustc.edu.cn/lncreg/">bioinformatics.ustc.edu.cn/lncreg/</a>	[145]
Linc2GO	Function inference	Provides comprehensive functional annotations for human lincRNA. <a href="http://www.bioinfo.tsinghua.edu.cn/~liuke/Linc2GO">http://www.bioinfo.tsinghua.edu.cn/~liuke/Linc2GO</a>	[146]
FARNA	Function annotation	Integrates ncRNA information related to expression, pathways and diseases in a large number of human tissues and primary cells. <a href="http://www.cbrc.kaust.edu.sa/farna/">www.cbrc.kaust.edu.sa/farna/</a>	[147]
ViRBase	Database	Provides the scientific community with a resource for efficient browsing and visualization of virus-host ncRNA-associated interactions and interaction networks in viral infection. <a href="http://www.rna-society.org/virbase">http://www.rna-society.org/virbase</a>	[148]

Tools	Type	Major Function/web link	References
LncRNA2Target	Database	Stores lncRNA-to-target genes. Provides a web interface for searching targets of a particular lncRNA or for the lncRNAs that target a particular gene. <a href="https://www.lncrna2target.org/">https://www.lncrna2target.org/</a>	[149]
Lncin	Function annotation	Identifies lncRNAs-associated modules from protein interaction networks and predicts the function of lncRNAs based on the protein functions in the modules. <a href="http://lncin.ym.edu.tw">lncin.ym.edu.tw</a>	[150]
NPInter	Function annotation	Integrates experimentally verified functional interactions between noncoding RNAs (excluding tRNAs and rRNAs) and other biomolecules (proteins, RNA and genomic DNA). <a href="http://www.bioinfo.org.cn/NPInter">www.bioinfo.org.cn/NPInter</a>	[151]
CPC	Coding potential assessment	Distinguishes between coding and noncoding RNA. Uses a Support Vector Machine-based classifier to assess the protein-coding potential of a transcript. <a href="http://cpc.cbi.pku.edu.cn/">cpc.cbi.pku.edu.cn/</a>	[122]
CNCI	Coding potential assessment	Distinguishes between protein-coding and non-coding sequences independent of known annotations. Applies to a variety of species without whole-genome sequence or with poorly annotated information. <a href="https://github.com/www-bioinfo-org/CNCI">https://github.com/www-bioinfo-org/CNCI</a>	[124]
CPAT	Coding potential assessment	Distinguishes between coding and noncoding RNA. Uses a logistic regression model to assess the protein coding potential. <a href="http://rna-cpat.sourceforge.net/">rna-cpat.sourceforge.net/</a>	[125]
FEELnc	LncRNA prediction	Derives an automatically computed cut-off so it maximizes the lncRNA prediction sensitivity and specificity. <a href="https://github.com/tderrien/FEELnc">https://github.com/tderrien/FEELnc</a>	[127]
PLEK	LncRNA prediction	Uses k-mer scheme and a support vector machine (SVM) algorithm to distinguish lncRNAs from mRNAs. <a href="http://www.ibiomedical.net/plek/">http://www.ibiomedical.net/plek/</a>	[126]

**Table 4.** Overview of tools for the analysis of lncRNA sequence data.

**3.3. Tools for identification of other non-coding RNA**

Currently, few tools have been developed for the identification of groups of ncRNAs other than miRNAs and lncRNAs. The popular tools for piRNA identification include ProTRAC [152], piClust [153], piRNAQuest [154], etc. (**Table 5**). proTRAC detects piRNA clusters based on a probabilistic analysis with assumption of a uniform distribution while piClust uses a density based clustering approach for the detection of piRNAs. piRNAQuest allows a search of the piRNome for silencers [154]. Another notable framework is SeqCluster [155], a python pipeline for the annotation and classification of non-miRNA small ncRNAs. The pipeline permits a

highly versatile and user-friendly interaction with data in order to easily classify small RNA sequences with putative functional importance [155]. For other small RNAs, ncPRO-seq [156] allows the discovery of unknown ncRNA or siRNA-coding regions from small RNA sequence data. DARIO [94] is a web-tool that allows annotation and detection of ncRNAs from various species but not livestock species. CoRAL [157] is a machine learning method that classifies ncRNAs by relying on biologically interpretable features. Several tools also have been developed for predicting circRNAs such as PredcircRNATool [158] and PredcircRNA [159] which apply a machine learning approach to distinguish circRNAs from other ncRNAs (**Table 5**).

Tools	Types	Main Features/web link	References
ProTRAC	piRNA prediction	Detects and analyses piRNA clusters based on quantifiable deviations from a hypothetical uniform distribution regarding the decisive piRNA cluster characteristics. <a href="https://sourceforge.net/projects/protrac/">https://sourceforge.net/projects/protrac/</a>	[152]
piClust	piRNA prediction	Finds piRNA clusters and transcripts from small RNA-seq data using a density based clustering approach. <a href="http://epigenomics.snu.ac.kr/piclustweb">http://epigenomics.snu.ac.kr/piclustweb</a>	[153]
piRNAQuest	piRNA database	Provides annotation of piRNAs based on their genomic location in gene, intron, intergenic, CDS, UTR, repeat elements, pseudogenes and syntenic regions. <a href="http://bicresources.jcbose.ac.in/zhumur/pirnaquest">bicresources.jcbose.ac.in/zhumur/pirnaquest</a>	[154]
SeqCluster	ncRNA classification	A framework python for the annotation and classification of the non-miRNA small RNA transcriptome. <a href="http://seqcluster.readthedocs.io/#">http://seqcluster.readthedocs.io/#</a>	[155]
ncPRO-seq	ncRNA discovery	Allows the discovery of unknown ncRNA- or siRNA-coding regions from sRNA sequence data. <a href="http://ncpro.curie.fr/">http://ncpro.curie.fr/</a> .	[156]
DARIO	ncRNA discovery	Allows annotation and detection of ncRNAs from various species but not livestock species. <a href="http://dario.bioinf.uni-leipzig.de/index.py">http://dario.bioinf.uni-leipzig.de/index.py</a>	[94]
CoRAL	ncRNA classification	A machine learning method that classifies ncRNA by relying on biologically interpretable features. <a href="http://wanglab.pcbi.upenn.edu/coral">http://wanglab.pcbi.upenn.edu/coral</a>	[157]
DASHR	Database	Stores human small ncRNAs: miRNAs, piRNAs, snRNAs, snoRNAs, scRNAs (small cytoplasmic RNAs), tRNAs, and rRNAs information. <a href="http://lisanwanglab.org/DASHR">lisanwanglab.org/DASHR</a>	[160]
Sno/scaRNABase	Database	A curated database for small nucleolar RNAs (snoRNAs) and small cajal body-specific RNAs (scaRNAs). <a href="http://gene.fudan.edu.cn/snoRNABase.nsf">gene.fudan.edu.cn/snoRNABase.nsf</a>	[161]

Tools	Types	Main Features/web link	References
snoRNA	Database	Contains over 1000 snoRNA sequences from Bacteria, Archaea, and Eukaryotes. <a href="http://evolveathome.com/snoRNA/snoRNA.php">http://evolveathome.com/snoRNA/snoRNA.php</a>	[162]
CircNet		Provides the following resources: (i) novel circRNAs, (ii) integrated miRNA-target networks, (iii) expression profiles of circRNA isoforms, (iv) genomic annotations of circRNA isoforms, and (v) sequences of circRNA isoforms. <a href="http://circnet.mbc.nctu.edu.tw">circnet.mbc.nctu.edu.tw</a>	[163]
PredicircRNATool	circRNA prediction	Uses a machine learning method for predicting circRNAs from those of non-circularized, expressed exons based on conformational and thermodynamic properties in the flanking introns. <a href="https://sourceforge.net/projects/predicircrnatool">https://sourceforge.net/projects/predicircrnatool</a>	[158]
circRNADb	circRNA database	Contains 32,914 human circular RNAs. <a href="http://reprod.njmu.edu.cn/circrnadb">http://reprod.njmu.edu.cn/circrnadb</a>	[164]
PredcircRNA	circRNA prediction	Applies a machine learning approach to predict circRNA. <a href="https://github.com/xypan1232/PredcircRNA">https://github.com/xypan1232/PredcircRNA</a>	[159]
CirsBase	Database	Provides scripts to identify known and novel circRNAs in sequence data. <a href="http://circbase.org">circbase.org</a>	[165]
Circ2Traits	Database	Contains a database of potential association of circular RNAs with diseases in human. <a href="http://gyanxet-beta.com/circdb">http://gyanxet-beta.com/circdb</a>	[166]
CircInteractome	Database	Provides a web tool for mapping (RNA Binding Proteins (RBP)- and miRNA-binding sites on human circRNAs. Allows to (i) identify potential circRNAs which can act as RBP sponges, (ii) design junction-spanning primers for specific detection of circRNAs of interest, (iii) design siRNAs for circRNA silencing, and (iv) identify potential internal ribosomal entry sites. <a href="https://circinteractome.nia.nih.gov">https://circinteractome.nia.nih.gov</a>	[167]
tRNADB	Database	Contains 12,000 tRNA genes from 577 species and 623 tRNA sequences from 104 species, provides various services such as graphical representations of tRNA secondary structures. <a href="http://trnad.bioinf.uni-leipzig.de">trnad.bioinf.uni-leipzig.de</a>	[168]

**Table 5.** Overview of tools and databases for sequence analysis of other small ncRNAs.

## 4. Tools for differential expression analysis of non-coding RNA

Various tools allow for the detection of genes (mRNA or ncRNA) differentially expressed (DE) between two or more conditions or states from sequence data. The major differences among tools are their implemented statistical methods, input and output file formats as well as filtering steps for DE analyses. Many tools such as DESeq [169], edgeR [170], NBPSseq [171], TSPM [172], baySeq [173], EBSeq [174], NOISeq [175], SAMseq [176] and ShrinkSeq [177] use count data as input file, while others like limma [178] and Cufflinks use transformed data or BAM files (the binary version of sequence alignment data) as input, respectively. Tools that use count data can be divided in to two groups; parametric (DESeq [169], edgeR [170], NBPSseq [171], TSPM [172], baySeq [173], EBSeq [174]) and non-parametric methods (NOISeq [175], SAMseq [176]). For parametric methods, most softwares (baySeq [173], DESeq [169], NBPSseq [171], edgeR [170], EBSeq [174] and NBPSseq) use a negative binomial model to account for over dispersion except ShrinkSeq which has two options for distribution, either negative binomial or a zero-inflated negative binomial distribution. These methods also implement different statistical test approaches; DESeq, edgeR and NBPSseq perform a classical hypothesis testing approach while baySeq, EBSeq and ShrinkSeq apply Bayesian methods. The comparison of methods and performances have been done and reviewed by many authors [29, 179–183]. In general, no single method performs well for all datasets. In a survey of performance of DE analyses methods, Conesa et al. [29] observed that limma package [178] performed well under many conditions. Many studies observed similar performances by DESeq and edgeR in ranking genes [29, 179–183]. However, DESeq is more conservative while edgeR is more liberal in controlling false discovery rate (FDR) [29]. Other tools such as SAMseq is better in controlling FDR while NOISeq is efficient in avoiding false positives [29].

## 5. Bioinformatics tools for target prediction and functional inference of non-coding RNA

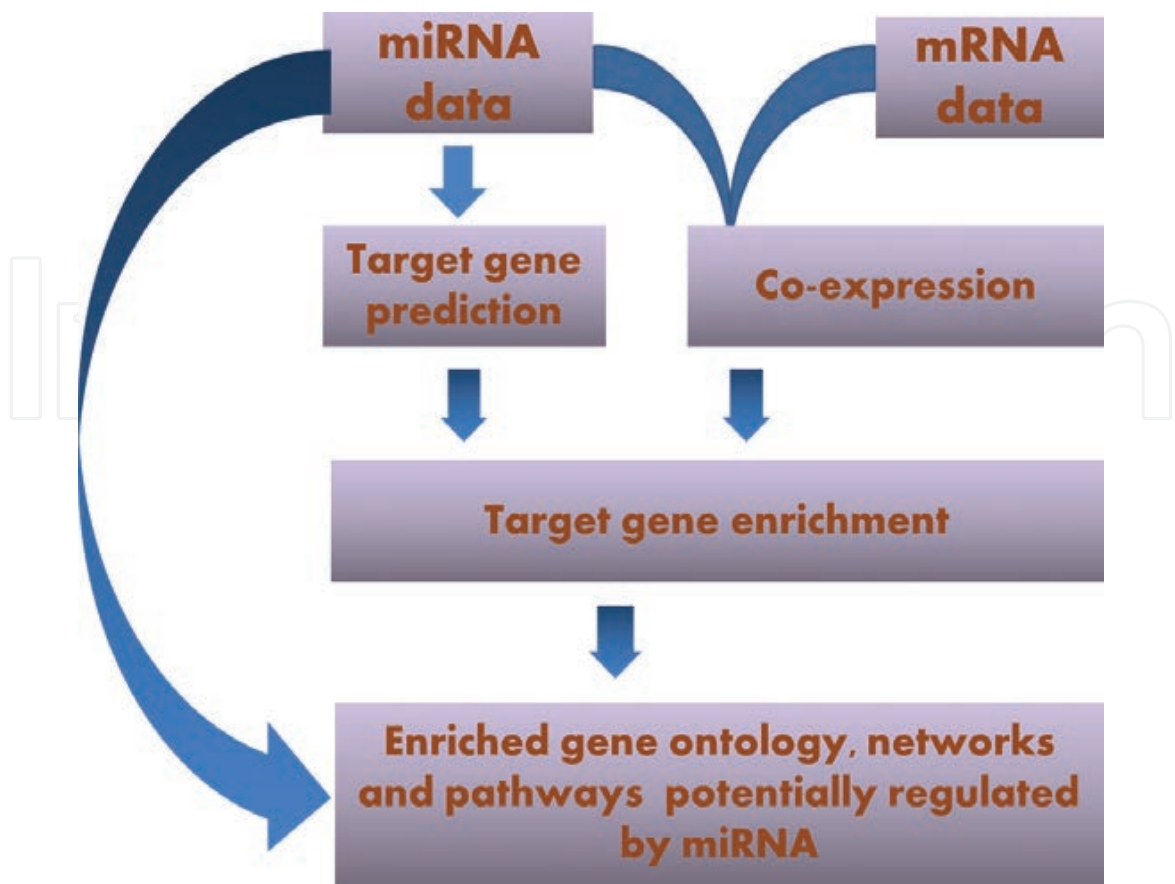
Following discovery and detection of important ncRNAs from RNA sequence data, the important next steps are to understand their regulatory roles. Since ncRNAs commonly act by interacting with target genes (mostly inhibit expression), various tools have been developed to predict their target genes and to infer their functions (**Tables 3 and 4**). A simple work flow for inferring the functions of miRNAs is shown in **Figure 4**.

### 5.1. Functional inference of miRNAs

#### 5.1.1. Bioinformatics tools for target prediction and functional inference of miRNAs

Inferring individual targets for a given miRNA can be done either by computational or experimental methods. Computational target prediction is coordinated in a sequence-specific manner and the target genes are normally predicted based on information derived from the





**Figure 4.** A simple work flow for inference of miRNA function.

potency of binding between miRNA and putative targets. Generally, the methods for computational prediction of miRNA targets can be grouped in single platforms such as TargetScan [95], PicTar [115], RNAhybrid [105] or multiple platforms such as miRwalk [116], TarBases [121], miRecords [117] as well as integrative platforms which include downstream analyses of putative target genes such as DIANA-microT-CDS [96], miRPathDB [184], etc. A collection of tools for miRNA target prediction are available at <https://omictools.com/mirna-target-prediction-category> and [https://tools4mirs.org/software/target\\_prediction/](https://tools4mirs.org/software/target_prediction/) [185] (**Table 3**). Among the prediction tools, the major differences in principles are in the algorithm applied and in filtering steps considering the secondary structure of the target mRNA (reviewed in [83, 115, 186]). Consequently, the specificity, sensitivity and accuracy of prediction are different among tools. Additionally, the performances of tools also differ based on the skills of the user (such as formatting of input and output, programming skills, web interface and so on). Taken together, all these factors affect popularity of tools [72, 187]. A word cloud plot of the popularity of tools based on their citation per year is shown in **Figure 5**.

*5.1.2. Popular single platforms for miRNA target prediction*

TargetScan can be accessed via the web interface or by running a perl script (local run) [95]. The software detects targets in the 3'UTR of protein-coding transcripts by base-pairing rules (seed complementarity) and predicts miRNAs for miRNA families instead of individual



binding sites data resulting from 12 target prediction programs (DIANA-microTv4.0, DIANA-microT-CDS, miRanda-rel2010, mirBridge, miRDB4.0, miRmap, miRNAMap, doRiNA, PicTar2, PITA, RNA22v2, RNAhybrid2.1 and Targetscan6.2) to build platforms of binding sites for the promoter, coding (5 prediction datasets), 5' and 3'UTR regions. It also contains experimentally verified miRNA-target interaction information collected via text-mining search and data from existing resources (miRTarBase, PhenomiR, miR2Disease and HMDD). MirRecords is a resource for animal miRNA-target interactions developed at the University of Minnesota [117]. MiRecords integrates predicted miRNA targets produced by 10 miRNA target prediction programs (DIANA-microTv4.0, miRanda-rel2010, miRDB4.0, PicTar2, PITA, RNAhybrid2.1, Targetscan6.2, miRTarget2, microinspector, NBmiRTar). It also contains information on experimentally validated miRNA targets obtained from the literature. mirDIP integrates 12 miRNA prediction datasets from miRNA prediction databases (DIANA-microTv4.0, miRanda-rel2010, miRDB4.0, PicTar2, PITA, RNAhybrid2.1, Targetscan6.2 and microCosm) allowing to customize miRNA target searches. multiMiR contains a collection of nearly 50 million records from 14 different databases [118]. It allows user-defined cut-offs for predicted binding strength to provide the most confident selection.

#### 5.1.4. Integrated tools for miRNA analysis

Various integrated tools as well as work flow for miRNA analysis have been developed to perform downstream analyses of putative target genes (e.g. gene ontology, pathways enrichments of target genes, etc.) such as MMIA [101], MAGIA [109] and miRconnX [119], to link miRNA to transcription factors or to analyze the effect of several miRNAs such as DIANA-mirExTra v2.0 [120] and TransMIR [114]. Typically, predicted target genes are used as input for functional enrichment to infer the potential functions of miRNAs. Furthermore, several tools are also used to correlate the expression levels of miRNAs with mRNA in a particular experiment to infer miRNA function such as miRnet [110], miRSystem [111] and DIANA-miRPath v3.0 [107]. Several tools have also been developed to directly link miRNAs to biological processes such as DMirNet [188], miRnet [110] and DIANA-miRPath v3.0 [107]. Many tools and resources have also been developed to link miRNAs to specific phenotypes/environments including diseases such as miRNAs in obsessive-compulsive disorder [189], autophagy in gerontology [190], epilepsy [191] and cancer [192]. Among the most popular integrated tools, DIANA-tools ([www.micorna.gr](http://www.micorna.gr)) covers a wide scope and research scenarios integrating several tools such as DIANA-microT-CDS, DIANA-TarBase v7.0, DIANA-miRGen v3.0, DIANA-miRPath v3.0, and DIANA-mirExTra v2.0. DIANA-microT-CDS uses different thresholds and meta-analysis followed by pathway enrichment to perform miRNA target prediction [96]. DIANA-TarBase is a manually curated target database with more than half a million miRNA-target interactions curated from published experiments performed with 356 different cell types from 24 species. DIANA-miRPath is an online software suite dedicated to the assessment of miRNA regulatory roles and the identification of controlled pathways [107]. DIANA-mirExTra performs combined differential expression analysis of mRNAs and miRNAs to uncover miRNAs and transcription factors that play important regulatory roles between two investigated state [193]. miRNet is an easy-to-use web-based tool for statistical analysis and functional interpretation of various datasets generated in miRNAs studies in

various species. Moreover, it also allows users to explore the results of miRNA-target interaction [110]. MMIA is a web tool for integration of miRNA and mRNA expression data with predicted miRNA target information for analyzing miRNA-associated phenotypes and biological functions by gene set enrichment analyses [101].

## 5.2. Functional inference of lncRNA

Compared to miRNAs, fewer bioinformatics tools have been developed for functional inference of lncRNAs. Several databases have been developed to curate computationally predicted and experimentally verified lncRNAs, such as LncRNAdb [194], GENCODE [137], lncRNAtor [7], lncRNome [195], NONCODE [135], lncRNAWiki [134], LncRNA2Function [143] and starBase v2.0 [196]. LncRNAdb was the first lncRNA database [194] and its updated version (LncRNAdb v2.0) integrates lncRNAs reported in livestock species (cattle, sheep, pig, horse and chicken) [131]. DeepBase database is an online platform for annotation and discovery of lncRNAs from RNA-seq data and it contains a large number of transcript entries for bovine (43,156) and chicken (47,004) lncRNAs. Other databases for livestock species are RNAcentral [197] which currently houses information from 23 ncRNA databases (<http://rnacentral.org/>, access March, 2017) but only contains a small number of lncRNAs from livestock species (cattle, pig, horse and chicken). NONCODE [135] contains lncRNAs for 16 species including cattle and chicken in the latest version. The first lncRNA database with a particular focus on domesticated animals was ALDB [136]. ALDB contains 12,103 pig lincRNAs (long intergenic non-coding RNA), 8923 chicken lincRNAs, and 8250 cow lincRNAs (<http://www.ibiomedical.net/aldb/>, access March, 2017). However, no comprehensive database currently covers available information on lncRNAs from livestock species, therefore the availability of a comprehensive tool will be valuable and helpful for subsequent genomic and functional annotation of lncRNAs and comparative interspecies analyses [198]. Inference of lncRNAs functions can also be done by connecting their expression patterns with specific cell types or biological processes to draw possible conclusions on their potential roles. lncRNAs can act in cis and/or trans manner to influence or interact with nearby or distant genes, respectively [2, 199]. For cis-regulation, the genomic location can be used as a guide for guilt-by-association analysis which allows global understanding of lncRNAs and protein coding genes that are tightly co-expressed and thus presumably co-regulated. Cis-relationships can foreseeably arise through complementary sequence motifs, tethering, blocking, and product-independent transcription [2]. For example, the human HOTTIP lncRNA is a cis-acting lncRNA expressed in the HOXA cluster that activates transcription of flanking genes [200]. The bioinformatics tools for cis-regulation prediction include ncFANs (<http://www.ebiomed.org/ncFANs>) [201] which uses a coding-non-coding gene co-expression network to infer lncRNA function.

## 6. Emerging platforms and technologies for understanding and using ncRNAs

Efficient and reliable techniques for accurate detection of genome information are important for productivity and health of livestock species [202]. The introduction of next generation



sequencing technologies has increased throughput studies of ncRNAs considerably. Consequently, studies on ncRNAs have contributed toward better understanding of disease resistance, productivity, breeding and meat quality in livestock species [203]. Although the numbers of detected ncRNA transcripts are increasing continuously, the ncRNAs identified and annotated in livestock species are still very scanty, compared with human data. Therefore, there is need to continue to explore the ncRNA transcriptome of livestock species [204]. The ability to explore and modify the genomes of livestock species could be beneficial in improving disease resistance, productivity, breeding capability as well as generation of new biomedical models [205].

Genome editing tools have emerged that allow efficient and precise genome manipulation of many organisms including livestock. The genome editing technique is built on engineered, programmable and highly specific nucleases that induce site-specific changes in the genomes of cellular organisms [206]. Subsequent cellular DNA repair processes generates desired insertions, deletions or substitutions at the loci of interest establishing linkages between genetic variations and biological phenotypes [207]. Presently, four artificially engineered nuclease systems have been developed for genome editing: meganucleases derived from microbial mobile elements, zinc finger nucleases (ZFNs) based on eukaryotic transcription factor DNA binding motif, transcription activator-like effector-based nucleases (TALEN) derived from a plant-invasive bacterial protein, and clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated protein 9 (Cas9) system [208]. Centromere and Promoter Factor 1 (Cpf1) is used as an alternative to Cas9 nuclease which requires only a single CRISPR RNA (crRNA) for targeting [209]. CRISPR/Cas9 is easily applicable and has developed really fast over the past years since only programmable RNA is required to generate sequence specificity [210].

CRISPR–Cas9 system is based on a bacterial CRISPR-Cas9 nuclease from *Streptococcus pyogenes* enabling inexpensive and high-throughput interrogation of gene function [211]. CRISPR-based screening can be used to study non-coding sequences, characterize enhancer elements and regulatory sequences crucial to elucidate the roles of ncRNA [212]. With the CRISPR–Cas9 system, the genome can be sliced at specific sites [213]. Genome editing techniques have been modified and used to alter the genomes of many organisms, thus offering opportunities for generation of genetically modified farm animals [214]. CRISPR offers the ability to target and study particular DNA sequences in the vast expanse of a genome [215]. There are two chief ingredients in the CRISPR–Cas9 system: a Cas9 enzyme that snips through DNA like a pair of molecular scissors, and a small RNA molecule that directs the scissors to a specific sequence of DNA to make the cut. The genome can be edited as desired at nearly any site if a template is provided [216].

In order to adapt this far-reaching application of gene-editing technology to agricultural improvement, various approaches have been applied to a number of livestock species. In pigs, direct cytoplasmic injection of Cas9 mRNA and single-guide RNA into zygotes generated biallelic knockout piglets [217]. The CRISPR-Cas9 system was used to generate gene-edited pigs protected from porcine reproductive and respiratory syndrome virus [218] and to genetically modify single blastocyst inducing indel mutations in a given gene locus [219]. Both TALEN and ZNF have been injected directly into pig zygotes to produce live genome edited pigs [220]. Similarly, the porcine myostatin (MSTN) gene, which functions as a negative regulator of muscle growth, was

disrupted using CRISPR/Cas9 system to efficiently generate biologically safe genetically modified pigs [221]. Similarly, zygote injection of TALEN mRNA targeting MSTN gene led to production of gene-edited cattle and sheep [205]

In cattle, the CRISPR/Cas9 system was successfully used to clone embryos that could be used to develop livestock transgenes for agricultural science [222]. Hornlessness was introduced into dairy cattle by genome editing and reproductive cloning providing the potential to improve the welfare of millions of cattle [223]. In the cattle industry, gene-edited calves have been produced with specified genetics by ovum pickup, *in vitro* fertilization and zygote microinjection (OPU-IVF-ZM). The CRISPR/Cas9 system has also been used efficiently to generate gene knock out sheep [224].

In livestock, CRISPR-Cas9 has been greatly enhanced by single-guide RNA generating site-specific DNA breaks through homology-directed repair and used for diverse applications, from disease modelling of individual loci to parallelized loss-of-function screens of thousands of regulatory elements [225]. Equally, bioinformatics designs for CRISPR deletions are now possible with a tool known as CRISPETa developed with efficient CRISPR deletion of an enhancer and exonic fragment of MALAT1, a lncRNA. CRISPETa can be used for single target regions or thousands of targets and has high-coverage library designs for entire classes of non-coding elements which can be adopted for use in livestock species [226]. CRISPR-Cas9 may be used with a gene drive incorporated with genome edit to investigate the control of any biological process and can be used to accelerate livestock breeding [225]. Gene drives can be constructed with the use of CRISPR-Cas9 tool that can favour the inheritance of edited alleles possible to modify a whole population [227]. In the DNA, a double strand break can be initiated by a gene drive during the copying process. Using the sequence of the chromosome containing the gene drive elements as a repair template, the DNA break could be repaired by cellular pathways such as homology-directed repair [228]. Editing the genomic DNA elements targeting non-coding regions is vital since silencing of ncRNA genes using RNA interference tools still presents major challenges. An improved vector system adapted to delete non-protein-coding regulatory elements; double excision CRISPR Knockout (DECKO) using two-step cloning to produce vectors (lentivirus) with two guide RNAs concurrently [229], has been used effectively to silenced five ncRNAs (miRNAs-miR21, miR29a and lncRNAs-UCA1 and MALAT1) [230]. The use of genome editing technologies will create novel viewpoints for enquiry to advance our knowledge on biological function of ncRNAs in livestock species and facilitate creating animals with precise alterations.

## 7. Conclusion and remarks

With the application of next generation sequencing technologies, the number of ncRNAs reported in livestock species has increased dramatically in the last 5 years. Various tools and pipelines have been introduced to make sense out of ncRNA sequence data. This chapter has provided a comprehensive overview of the current and emerging tools and methods for generating and analyzing ncRNA (miRNA, lncRNA as well as other small ncRNAs) sequence



data (transcriptome) with special emphases on the tools that can be applied to livestock species. While bioinformatics tools for miRNA analyses are quite mature, there is a general lack of comprehensive bioinformatics tools for lncRNA and other small ncRNAs. It is our belief that comprehensive “omics” databases that integrate existing and future ncRNA transcriptome databases in the framework of livestock species will contribute towards elucidation of the ambiguity surrounding RNA sequence data. Moreover, given the fact that several emerging platforms (such as genome editing tools) for understanding ncRNAs have been introduced recently, these tools certainly bring great opportunities for broader and also deeper exploration of ncRNA functions. In addition, meticulous *in silico* prediction and careful interpretation of results are critical when handling ncRNA sequence data. Finally, wet-lab validation of the results of transcriptome data will be vital to confirm the functions of ncRNAs in livestock species.

## Acknowledgements

We acknowledge financial support from Agriculture and Agri-Food Canada.

## Author details

Duy N. Do<sup>1,2</sup>, Pier-Luc Dudemaine<sup>1</sup>, Bridget Fomenky<sup>1,3</sup> and Eveline M. Ibeagha-Awemu<sup>1\*</sup>

\*Address all correspondence to: eveline.ibeagha-awemu@agr.gc.ca

1 Agriculture and Agri-Food Canada, Sherbrooke Research and Development Centre, Sherbrooke, Quebec, Canada

2 Department of Animal Science, McGill University, Ste-Anne-de Bellevue, Quebec, Canada

3 Département des Sciences Animales, Université Laval, Québec, QC, Canada

## References

- [1] Mercer TR, Wilhelm D, Dinger ME, Solda G, Korbie DJ, Glazov EA, Truong V, Schwenke M, Simons C, Matthaei KI. Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Research*. 2011;**39**(6):2393-2403
- [2] Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*. 2009;**10**(3):155-159
- [3] Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA*. 2010;**16**(4):673-695

- [4] Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & Development*. 2009;**23**(22):2639-2649
- [5] Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. A human snoRNA with microRNA-like functions. *Molecular Cell*. 2008;**32**(4): 519-528
- [6] Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, Mattick JS. Small RNAs derived from snoRNAs. *RNA*. 2009;**15**(7):1233-1240
- [7] Matera AG, Terns RM, Terns MP. Non-coding RNAs: Lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews Molecular Cell Biology*. 2007;**8**(3):209-220
- [8] Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics*. 2014;**15**(1):7-21
- [9] Stefani G, Slack FJ. Small non-coding RNAs in animal development. *Nature Reviews Molecular Cell Biology*. 2008;**9**(3):219-230
- [10] Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: Regulators of disease. *The Journal of Pathology*. 2010;**220**(2):126-139
- [11] Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 2009;**10**(1):57-63
- [12] Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nature reviews Genetics*. 2016;**17**(6):333-351
- [13] Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;**270**(5235):467-470
- [14] Tarca AL, Romero R, Draghici S. Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology*. 2006;**195**(2):373-388
- [15] Kroll KM, Barkema GT, Carlon E. Modeling background intensity in DNA microarrays. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*. 2008;**77**(6 Pt 1):061915
- [16] Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association*. 2004;**99**(468):909-917
- [17] Schreiber K, Csaba G, Haslbeck M, Zimmer R. Alternative splicing in next generation sequencing data of *Saccharomyces cerevisiae*. *PLoS One*. 2015;**10**(10):e0140487
- [18] Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. 2011;**27**(17):2325-2329
- [19] Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. *American Journal of Human Genetics*. 2013;**93**(4):641-651

- [20] Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*. 2008;**5**(7):613-619
- [21] Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, McDonald H, Varhol R, Jones S, Marra M. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques*. 2008;**45**(1):81-94
- [22] Wang Y, Xue S, Liu X, Liu H, Hu T, Qiu X, Zhang J, Lei M. Analyses of Long Non-Coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. *Scientific Report*. 2016;**6**:20238
- [23] Bottomly D, Walter NA, Hunter JE, Darakjian P, Kawane S, Buck KJ, Searles RP, Mooney M, McWeeney SK, Hitzemann R. Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. *PLoS One*. 2011;**6**(3):e17820
- [24] Sirbu A, Kerr G, Crane M, Ruskin HJ. RNA-Seq vs dual- and single-channel microarray data: Sensitivity analysis for differential expression and clustering. *PLoS One*. 2012;**7**(12):e50986
- [25] Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, Menzel C, Chen W, Li Y, Zeng R et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics*. 2009;**10**:161
- [26] Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*. 2016;**17**(1):239
- [27] Chu Y, Corey DR. RNA sequencing: Platform selection, experimental design, and data interpretation. *Nucleic Acid Therapeutics*. 2012;**22**(4):271-274
- [28] Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*. 2011;**8**(6):469-477
- [29] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X. A survey of best practices for RNA-seq data analysis. *Genome Biology*. 2016;**17**(1):13
- [30] Liu Y, Ferguson JF, Xue C, Silverman IM, Gregory B, Reilly MP, Li M. Evaluating the impact of sequencing depth on transcriptome profiling in human adipose. *PLoS One*. 2013;**8**(6):e66883
- [31] Liu Y, Zhou J, White KP. RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*. 2014;**30**(3):301-304
- [32] Podolska A, Kaczkowski B, Litman T, Fredholm M, Cirera S. How the RNA isolation method can affect microRNA microarray results. *Acta Biochimica Polonica*. 2011;**58**(4):535-540
- [33] Campbell JD, Liu G, Luo L, Xiao J, Gerrein J, Juan-Guardela B, Tedrow J, Alekseyev YO, Yang IV, Correll M et al. Assessment of microRNA differential expression and detection in multiplexed small RNA sequencing data. *RNA*. 2015;**21**(2):164-171

- [34] Metpally RP, Nasser S, Malenica I, Courtright A, Carlson E, Ghaffari L, Villa S, Tembe W, Van Keuren-Jensen K. Comparison of analysis tools for miRNA high throughput sequencing using nerve crush as a model. *Frontiers in Genetics*. 2013;**4**:20
- [35] Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- [36] Patel RK, Jain M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PloS One*. 2012;**7**(2):e30619
- [37] Williams CR, Baccarella A, Parrish JZ, Kim CC. Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics*. 2016;**17**:103
- [38] Chen C, Khaleel SS, Huang H, Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code for Biology and Medicine*. 2014;**9**:8-8
- [39] Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;**30**(15):2114-2120
- [40] Gordon A, Hannon G. Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished). [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/); 2010
- [41] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011;**17**(1): Next Generation Sequencing Data Analysis
- [42] Mielczarek M, Szyda J. Review of alignment and SNP calling algorithms for next-generation sequencing data. *Journal of Applied Genetics*. 2016;**57**(1):71-79
- [43] Shang J, Zhu F, Vongsangnak W, Tang Y, Zhang W, Shen B. Evaluation and comparison of multiple aligners for next-generation sequencing data analysis. *BioMed Research International*. 2014;**2014**:309650
- [44] Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;**25**(9):1105-1111
- [45] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;**29**(1):15-21
- [46] Langmead B. Aligning short sequencing reads with Bowtie. *Current Protocols in Bioinformatics*. 2010, Chapter 11:Unit 11 17
- [47] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*. 2015;**33**(3):290-295
- [48] Yang C, Wu PY, Tong L, Phan JH, Wang MD. The impact of RNA-seq aligners on gene expression estimation. *ACM BCB*. 2015;**2015**:462-471
- [49] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*. 2013;**8**(8):1494-1512

- [50] Li YL, Weng JC, Hsiao CC, Chou MT, Tseng CW, Hung JH. PEAT: An intelligent and efficient paired-end sequencing adapter trimming algorithm. *BMC Bioinformatics*. 2015;**16**(Suppl 1):S2
- [51] Wu Z, Wang X, Zhang X. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics*. 2011;**27**(4):502-508
- [52] Jiang H, Lei R, Ding SW, Zhu S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*. 2014;**15**:182
- [53] Criscuolo A, Brisse S. AlienTrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics*. 2013;**102**(5-6):500-506
- [54] O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. NxTrim: Optimized trimming of Illumina mate pair reads. *Bioinformatics*. 2015;**31**(12):2035-2037
- [55] Sturm M, Schroeder C, Bauer P. SeqPurge: Highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics*. 2016;**17**:208
- [56] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012;**9**(4):357-359
- [57] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 2009;**10**(3):R25
- [58] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;**26**(5):589-595
- [59] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. 2013;**14**(4):R36
- [60] McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, Vanderpool CK, Tjaden B. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research*. 2013;**41**(14):e140
- [61] Au KF, Jiang H, Lin L, Xing Y, Wong WH. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Research*. 2010;**38**(14):4570-4578
- [62] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M et al. De novo transcript sequence reconstruction from RNA-Seq: Reference generation and analysis with Trinity. *Nature Protocols*. 2013;**8**(8):1494-1512. DOI: 10.1038/nprot.2013.1084
- [63] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010;**28**(5):511-515



- [64] Mezlini AM, Smith EJ, Fiume M, Buske O, Savich GL, Shah S, Aparicio S, Chiang DY, Goldenberg A, Brudno M. iReckon: Simultaneous isoform discovery and abundance estimation from RNA-seq data. *Genome Research*. 2013;**23**(3):519-529
- [65] Liu NY, Xu W, Papanicolaou A, Dong SL, Anderson A. Identification and characterization of three chemosensory receptor families in the cotton bollworm *Helicoverpa armigera*. *BMC Genomics*. 2014;**15**:597
- [66] Tsoi LC, Iyer MK, Stuart PE, Swindell WR, Gudjonsson JE, Tejasvi T, Sarkar MK, Li B, Ding J, Voorhees JJ et al. Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. *Genome Biology*. 2015;**16**:24
- [67] Amin V, Harris RA, Onuchic V, Jackson AR, Charnecki T, Paithankar S, Lakshmi Subramanian S, Riehle K, Coarfa C, Milosavljevic A. Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs. *Nature Communications*. 2015;**6**:6370
- [68] Koufariotis LT, Chen Y-PP, Chamberlain A, Vander Jagt C, Hayes BJ. A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLoS One*. 2015;**10**(10):e0141225
- [69] Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*. 2012;**40**(1):37-52
- [70] Hackenberg M, Rodriguez-Ezpeleta N, Aransay AM. miRanalyzer: An update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Research*. 2011;**39**(Web Server issue):W132-W138
- [71] Wu J, Liu Q, Wang X, Zheng J, Wang T, You M, Sheng Sun Z, Shi Q. mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. *RNA Biology*. 2013;**10**(7):1087-1092
- [72] Akhtar MM, Micolucci L, Islam MS, Olivieri F, Procopio AD: Bioinformatic tools for microRNA dissection. *Nucleic Acids Research*. 2016;**44**(1):24-44
- [73] Shukla V, Varghese VK, Kabekkodu SP, Mallya S, Satyamoorthy K. A compilation of Web-based research tools for miRNA analysis. *Briefings in Functional Genomics*. 2017. <https://doi.org/10.1093/bfpg/elw042>
- [74] Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*. 2012;**40**(1):37-52
- [75] Hackenberg M, Sturm M, Langenberger D, Falcon-Perez JM, Aransay AM. miRanalyzer: A microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Research*. 2009;**37**(suppl 2):W68-W76
- [76] Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, Folkes L, Schwach F, Dalmay T, Moulton V. The UEA sRNA workbench: A suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*. 2012;**28**(15):2059-2061



- [77] Rueda A, Barturen G, Lebrón R, Gómez-Martín C, Alganza Á, Oliver JL, Hackenberg M. sRNAtoolbox: An integrated collection of small RNA research tools. *Nucleic Acids Research*. 2015;**43**(W1):W467-W473
- [78] Pantano L, Estivill X, Martí E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Research*. 2010;**38**(5):e34-e34
- [79] Jiang P, Wu H, Wang W, Ma W, Sun X, Lu Z. MiPred: Classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Research*. 2007;**35**(Suppl 2):W339-W344
- [80] Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, Brownstein MJ, Tuschl T, van Nimwegen E, Zavolan M. Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics*. 2005;**6**(1):267
- [81] Mathelier A, Carbone A. MIRENA: Finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics*. 2010;**26**(18):2226-2234
- [82] Gomes CPC, Cho J-H, Hood L, Franco OL, Pereira RW, Wang K. A review of computational tools in microRNA discovery. *Frontiers in Genetics*. 2013;**4**:81
- [83] Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB. Common features of microRNA target prediction tools. *Frontiers in Genetics*. 2014;**5**:23
- [84] Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE. Mammalian microRNAs: Experimental evaluation of novel and previously annotated genes. *Genes & Development*. 2010;**24**(10):992-1009
- [85] Li Z, Liu H, Jin X, Lo L, Liu J. Expression profiles of microRNAs from lactating and non-lactating bovine mammary glands and identification of miRNA related to lactation. *BMC Genomics*. 2012;**13**(1):731
- [86] Kozomara A, Griffiths-Jones S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*. 2014;**42**(D1):D68-D73
- [87] Markham NR, Zuker M. UNAFold: Software for nucleic acid folding and hybridization. *Bioinformatics: Structure, Function and Applications*. 2008:3-31
- [88] Peng J, Zhao J-S, Shen Y-F, Mao H-G, Xu N-Y. MicroRNA expression profiling of lactating mammary gland in divergent phenotype swine breeds. *International Journal of Molecular Sciences*. 2015;**16**(1):1448-1465
- [89] Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The vienna RNA websuite. *Nucleic Acids Research*. 2008;**36**(suppl 2):W70-W74
- [90] Li R, Dudemaine P-L, Zhao X, Lei C, Ibeagha-Awemu EM. Comparative analysis of the miRNome of bovine milk fat, whey and cells. *PloS One*. 2016;**11**(4):e0154129
- [91] Schroeder DI, Jayashankar K, Douglas KC, Thirkill TL, York D, Dickinson PJ, Williams LE, Samollow PB, Ross PJ, Bannasch DL. Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. *PLoS Genetics*. 2015;**11**:e1005442

- [92] Do DN, Li R, Dudemaine P-L, Ibeagha-Awemu EM. MicroRNA roles in signalling during lactation: An insight from differential expression, time course and pathway analyses of deep sequence data. *Scientific Reports*. 2017;**7**:44605
- [93] Wang W-C, Lin F-M, Chang W-C, Lin K-Y, Huang H-D, Lin N-S. miRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics*. 2009;**10**(1):328
- [94] Fasold M, Langenberger D, Binder H, Stadler PF, Hoffmann S. DARIO: A ncRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Research*. 2011;**39**(Web Server issue):W112-W117: gkr357
- [95] Lewis BP, Shih I-h, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell*. 2003;**115**(7):787-798
- [96] Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG. DIANA-microT web server v5. 0: Service integration into miRNA functional analysis workflows. *Nucleic Acids Research*. 2013;**41**(W1):W169-W173
- [97] Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in *Drosophila*. *Genome Biology*. 2003;**5**(1):R1
- [98] Wong N, Wang X. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Research*. 2014;**43**(D1):D146-D152. gku1104
- [99] Hsu JBK, Chiu CM, Hsu SD, Huang WY, Chien CH, Lee TY, Huang HD. miRTar: An integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics*. 2011;**12**(1):300
- [100] Hammell M, Long D, Zhang L, Lee A, Carmack CS, Han M, Ding Y, Ambros V. mir-WIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nature Methods*. 2008;**5**(9):813-819
- [101] Nam S, Li M, Choi K, Balch C, Kim S, Nephew KP. MicroRNA and mRNA integrated analysis (MMIA): A web tool for examining biological functions of microRNA expression. *Nucleic Acids Research*. 2009;**37**(suppl 2):W356-W362
- [102] Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nature Genetics*. 2007;**39**(10):1278-1284
- [103] Dai X, Zhao PX. psRNATarget: A plant small RNA target analysis server. *Nucleic Acids Research*. 2011;**39**(suppl 2):W155-W159
- [104] Miranda KC, Huynh T, Tay Y, Ang Y-S, Tam W-L, Thomson AM, Lim B, Rigoutsos I. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*. 2006;**126**(6):1203-1217
- [105] Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research*. 2006;**34**(Suppl 2):W451-W454

- [106] Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*. 2007;**13**(11): 1894-1910
- [107] Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG. DIANA-miRPath v3.0: Deciphering microRNA function with experimental support. *Nucleic Acids Research*. 2015;**43**(W1):W460-W466
- [108] Nam S, Kim B, Shin S, Lee S. miRGator: An integrated system for functional annotation of microRNAs. *Nucleic Acids Research*. 2008;**36**(suppl 1):D159-D164
- [109] Sales G, Coppe A, Bisognin A, Biasiolo M, Bortoluzzi S, Romualdi C. MAGIA, a web-based tool for miRNA and genes integrated analysis. *Nucleic Acids Research*. 2010;**38**(Web Server issue):W352-W359. gkq423
- [110] Fan Y, Siklenka K, Arora SK, Ribeiro P, Kimmins S, Xia J. miRNet-dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Research*. 2016;**44**(W1):W135-W141
- [111] Lu TP, Lee CY, Tsai MH, Chiu YC, Hsiao CK, Lai LC, Chuang EY. miRSystem: An integrated system for characterizing enriched functions and pathways of microRNA targets. *PloS One*. 2012;**7**(8):e42390
- [112] Hsu SD, Chu CH, Tsou AP, Chen SJ, Chen HC, Hsu PWC, Wong YH, Chen YH, Chen GH, Huang HD. miRNAMap 2.0: Genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Research*. 2008;**36**(Suppl 1):D165-D169
- [113] Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM. miRTarBase: A database curates experimentally validated microRNA–target interactions. *Nucleic Acids Research*. 2010;**39**(Database issue):D163-D169. gkq1107
- [114] Wang J, Lu M, Qiu C, Cui Q. TransmiR: A transcription factor–microRNA regulation database. *Nucleic Acids Research*. 2010;**38**(suppl 1):D119-D122
- [115] Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, Da Piedade I, Gunsalus KC, Stoffel M. Combinatorial microRNA target predictions. *Nature Genetics*. 2005;**37**(5):495-500
- [116] Dweep H, Sticht C, Pandey P, Gretz N. miRWalk–database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes. *Journal of Biomedical Informatics*. 2011;**44**(5):839-847
- [117] Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: An integrated resource for microRNA–target interactions. *Nucleic Acids Research*. 2009;**37**(suppl 1):D105-D110
- [118] Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, Mahaffey S, Rossi S, Calin GA, Bemis L. The multiMiR R package and database: Integration of microRNA–target interactions along with their disease and drug associations. *Nucleic Acids Research*. 2014;**42**(17):e133-e133

- [119] Huang GT, Athanassiou C, Benos PV. mirConnX: Condition-specific mRNA-microRNA network integrator. *Nucleic Acids Research*. 2011;**39**(suppl 2):W416-W423
- [120] Vlachos IS, Vergoulis T, Paraskevopoulou MD, Lykokanellos F, Georgakilas G, Georgiou P, Chatzopoulos S, Karagkouni D, Christodoulou F, Dalamagas T. DIANA-mirExTra v2.0: Uncovering microRNAs and transcription factors with crucial roles in NGS expression data. *Nucleic Acids Research*. 2016;**44**(Web Server issue):W128-W134. gkw455
- [121] Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, Gerangelos S, Koziris N, Dalamagas T, Hatzigeorgiou AG. TarBase 6.0: Capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Research*. 2012;**40**(D1):D222-D229
- [122] Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*. 2007;**35**(suppl 2):W345-W349
- [123] Lin MF, Jungreis I, Kellis M. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;**27**(13):i275-i282
- [124] Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research*. 2013;**41**(17):e166-e166
- [125] Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Research*. 2013;**41**(6):e74-e74
- [126] Li A, Zhang J, Zhou Z. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics*. 2014;**15**(1):311
- [127] Wucher V, Legeai F, Hedan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H. FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Research*. 2017;**45**(8):e57. gkw1306
- [128] Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, Casas E, Cheng HH, Clarke L, Couldrey C et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biology*. 2015;**16**(1):57
- [129] Yang J-H, Li J-H, Jiang S, Zhou H, Qu L-H. ChIPBase: A database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Research*. 2013;**41**(D1):D177-D187
- [130] Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P. LNCipedia: A database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Research*. 2013;**41**(D1):D246-D251
- [131] Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME. lncRNADB v2.0: Expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Research*. 2014;**43**(Database issue):D168-D173. gku988



- [132] Xu J, Bai J, Zhang X, Lv Y, Gong Y, Liu L, Zhao H, Yu F, Ping Y, Zhang G. A comprehensive overview of lncRNA annotation resources. *Briefings in bioinformatics*. 2016;**18**(2):236-249. bbw015
- [133] Gong J, Liu W, Zhang J, Miao X, Guo A-Y. lncRNASNP: A database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Research*. 2015;**43**(D1):D181-D186
- [134] Ma L, Li A, Zou D, Xu X, Xia L, Yu J, Bajic VB, Zhang Z. lncRNAWiki: Harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Research*. 2014;**43**(Database issue):D187-D192. gku1167
- [135] Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y. NONCODEv4: Exploring the world of long non-coding RNA genes. *Nucleic Acids Research*. 2014;**42**(D1):D98-D103
- [136] Cao J, Wei C, Liu D, Wang H, Wu M, Xie Z, Capellini TD, Zhang L, Zhao F, Li L. DNA methylation Landscape of body size variation in sheep. *Scientific Reports*. 2015;**5**
- [137] Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*. 2012;**22**(9):1775-1789
- [138] Wu D, Huang Y, Kang J, Li K, Bi X, Zhang T, Jin N, Hu Y, Tan P, Zhang L. ncRDeathDB: A comprehensive bioinformatics resource for deciphering network organization of the ncRNA-mediated cell death system. *Autophagy*. 2015;**11**(10):1917-1926
- [139] Chen X, Hao Y, Cui Y, Fan Z, He S, Luo J, Chen R. lncVar: A database of genetic variation associated with long non-coding genes. *Bioinformatics*. 2017;**33**(1):112-118
- [140] Denisenko E, Ho D, Tamgue O, Ozturk M, Suzuki H, Brombacher F, Guler R, Schmeier S. IRNdb: The database of immunologically relevant non-coding RNAs. *Database*. 2016;**2016**. baw138
- [141] Hou M, Tang X, Tian F, Shi F, Liu F, Gao G. AnnoLnc: A web server for systematically annotating novel human lncRNAs. *BMC Genomics*. 2016;**17**(1):931
- [142] He S, Zhang H, Liu H, Zhu H. LongTarget: A tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. *Bioinformatics*. 2015;**31**(2):178-186
- [143] Jiang Q, Ma R, Wang J, Wu X, Jin S, Peng J, Tan R, Zhang T, Li Y, Wang Y. lncRNA2Function: A comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data. *BMC Genomics*. 2015;**16**(3):S2
- [144] Zhao Z, Bai J, Wu A, Wang Y, Zhang J, Wang Z, Li Y, Xu J, Li X. Co-LncRNA: Investigating the lncRNA combinatorial effects in GO annotations and KEGG pathways based on human RNA-Seq data. *Database*. 2015;**2015**. bav082
- [145] Zhou Z, Shen Y, Khan MR, Li A. lncReg: A reference resource for lncRNA-associated regulatory networks. *Database*. 2015;**2015**. bav083



- [146] Liu K, Yan Z, Li Y, Sun Z. Linc2GO: A human LincRNA function annotation resource based on ceRNA hypothesis. *Bioinformatics*. 2013;**29**(17):2221-2222
- [147] Alam T, Uludag M, Essack M, Salhi A, Ashoor H, Hanks JB, Kapfer C, Mineta K, Gojobori T, Bajic VB. FARNA: Knowledgebase of inferred functions of non-coding RNA transcripts. *Nucleic Acids Research*. 2016;**45**(5):2838-2848. gkw973
- [148] Li Y, Wang C, Miao Z, Bi X, Wu D, Jin N, Wang L, Wu H, Qian K, Li C. ViRBase: A resource for virus–host ncRNA-associated interactions. *Nucleic Acids Research*. 2014;**43**(Database issue):D578-D582. gku903
- [149] Jiang Q, Wang J, Wu X, Ma R, Zhang T, Jin S, Han Z, Tan R, Peng J, Liu G. LncRNA2Target: A database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Research*. 2015;**43**(D1):D193-D196
- [150] Wu CH, Hsu CL, Lu PC, Lin WC, Juan HF, Huang HC. Identification of lncRNA functions in lung cancer based on associated protein-protein interaction modules. *Scientific Reports*. 2016;**6**:35959
- [151] Wu T, Wang J, Liu C, Zhang Y, Shi B, Zhu X, Zhang Z, Skogerbø G, Chen L, Lu H. NPInter: The noncoding RNAs and protein related biomacromolecules interaction database. *Nucleic Acids Research*. 2006;**34**(suppl 1):D150-D152
- [152] Rosenkranz D, Zischler H. proTRAC-a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*. 2012;**13**(1):5
- [153] Jung I, Park JC, Kim S. piClust: A density based piRNA clustering algorithm. *Computational Biology and Chemistry*. 2014;**50**:60-67
- [154] Sarkar A, Maji RK, Saha S, Ghosh Z. piRNAQuest: Searching the piRNAome for silencers. *BMC Genomics*. 2014;**15**(1):555
- [155] Pantano L, Estivill X, Martí E. A non-biased framework for the annotation and classification of the non-miRNA small RNA transcriptome. *Bioinformatics*. 2011;**27**(22):3202-3203
- [156] Chen C-J, Servant N, Toedling J, Sarazin A, Marchais A, Duvernois-Berthet E, Cognat V, Colot V, Voinnet O, Heard E et al. ncPRO-seq: A tool for annotation and profiling of ncRNAs in sRNA-seq data. *Bioinformatics*. 2012;**28**(23):3147-3149
- [157] Leung YY, Ryvkin P, Ungar LH, Gregory BD, Wang L-S. CoRAL: Predicting non-coding RNAs from small RNA-sequencing data. *Nucleic Acids Research*. 2013;**41**(14):e137. gkt426
- [158] Liu Z, Han J, Lv H, Liu J, Liu R. Computational identification of circular RNAs based on conformational and thermodynamic properties in the flanking introns. *Computational Biology and Chemistry*. 2016;**61**:221-225
- [159] Pan X, Xiong K. PredcircRNA: Computational classification of circular RNA from other long non-coding RNA using hybrid features. *Molecular Biosystems*. 2015;**11**(8):2219-2226
- [160] Leung YY, Kuksa PP, Amlie-Wolf A, Valladares O, Ungar LH, Kannan S, Gregory BD, Wang LS. DASHR: Database of small human noncoding RNAs. *Nucleic Acids Research*. 2015;**44**(D1):D216-D222. gkv1188

- [161] Xie J, Zhang M, Zhou T, Hua X, Tang L, Wu W. Sno/scaRNAbase: A curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Research*. 2007;**35**(suppl 1):D183-D187
- [162] Ellis JC, Brown DD, Brown JW. The small nucleolar ribonucleoprotein (snoRNP) database. *RNA*. 2010;**16**(4):664-666
- [163] Liu Y-C, Li J-R, Sun C-H, Andrews E, Chao R-F, Lin F-M, Weng S-L, Hsu S-D, Huang C-C, Cheng C. CircNet: A database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Research*. 2015;**44**(D1):D209-D215. gkv940
- [164] Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Scientific Reports*. 2016;**6**:34985
- [165] Glazar P, Papavasileiou P, Rajewsky N. circBase: A database for circular RNAs. *RNA*. 2014;**20**(11):1666-1670
- [166] Ghosal S, Das S, Sen R, Basak P, Chakrabarti J. Circ2Traits: A comprehensive database for circular RNA potentially associated with disease and traits. *Frontiers in Genetics*. 2013;**4**:283
- [167] Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biology*. 2016;**13**(1):34-42
- [168] Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J. tRNADB 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research*. 2009;**37** (Database issue):D159-D162
- [169] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology*. 2010;**11**
- [170] Robinson MD, McCarthy DJ, Smyth GK. EdgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;**26**(1):139-40
- [171] Di Y, Schafer DW, Cumbie JS, Chang JH. The NBP negative binomial model for assessing differential gene expression from RNA-seq. *Statistical Applications in Genetics and Molecular Biology*. 2011;**10**(1): 1-28
- [172] Auer PL, Doerge RW. A two-stage Poisson model for testing RNA-seq data. *Statistical Applications in Genetics and Molecular Biology*. 2011;**10**(1):1-26
- [173] Hardcastle TJ, Kelly KA. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinforma*. 2010;**11**:442
- [174] Leng, Ning, John A. Dawson, James A. Thomson, Victor Ruotti, Anna I. Rissman, Bart MG Smits, Jill D. Haag, Michael N. Gould, Ron M. Stewart, and Christina Kendzierski. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*. 2013;**29**(8):1035-1043
- [175] Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: A matter of depth. *Genome Research*. 2011;**21**(12):2213-2223

- [176] Li J, Tibshirani R. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-seq data. *Statistical Methods in Medical Research*. 2013;**22**(5):519-536
- [177] Van de Wiel MA, Leday GGR, Pardo L, Rue H, Van der Vaart AW, Van Wieringen WN. Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors. *Biostatistics*. 2012;**14**(1):113-128
- [178] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015;**43**(7):e47. gkv007
- [179] Sonesson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*. 2013;**14**(1):91
- [180] Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics*. 2015;**16**(1):59-70
- [181] Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, Robinson GJ, Lundberg AE, Bartlett PF, Wray NR. A comparative study of techniques for differential expression analysis on RNA-Seq data. *PloS One*. 2014;**9**(8):e103207
- [182] Robles JA, Qureshi SE, Stephen SJ, Wilson SR, Burden CJ, Taylor JM. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics*. 2012;**13**(1):484
- [183] Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*. 2013;**14**(9):3158
- [184] Backes C, Kehl T, Stöckel D, Fehlmann T, Schneider L, Meese E, Lenhof H-P, Keller A. miRPathDB: A new dictionary on microRNAs and target pathways. *Nucleic Acids Research*. 2017;**45**(D1):D90-D96
- [185] Lukasik A, Wójcikowski M, Zielenkiewicz P. Tools4miRs—one place to gather all the tools for miRNA analysis. *Bioinformatics*. 2016;**32**(17):2722-2724
- [186] Rajewsky N. microRNA target predictions in animals. *Nature Genetics*. 2006; **38**:S8-S13
- [187] Moore AC, Winkler JS, Tseng TT. Bioinformatics resources for microRNA discovery. *Biomarker Insights*. 2015;**10**(Suppl 4):53
- [188] Lee M, Lee H. DMirNet: Inferring direct microRNA-mRNA association networks. *BMC Systems Biology*. 2016;**10**(5):51
- [189] Privitera AP, Distefano R, Wefer HA, Ferro A, Pulvirenti A, Giugno R. OCDB: A database collecting genes, miRNAs and drugs for obsessive-compulsive disorder. *Database: The Journal of Biological Databases and Curation*. 2015;**2015**. bav069
- [190] Zhang L, Xie T, Tian M, Li J, Song S, Ouyang L, Liu B, Cai H. GAMDB: A web resource to connect microRNAs with autophagy in gerontology. *Cell Proliferation*. 2016;**49**(2):246-251

- [191] Mooney C, Becker BA, Raoof R, Henshall DC. EpimiRBase: A comprehensive database of microRNA-epilepsy associations. *Bioinformatics*. 2016;**32**(9):1436-1438
- [192] Dong L, Luo M, Wang F, Zhang J, Li T, Yu J. TUMIR: An experimentally supported database of microRNA deregulation in various cancers. *Journal of Clinical Bioinformatics*. 2013;**3**(1):7
- [193] Iftikhar H, Schultzhaus JN, Bennett CJ, Carney GE. The in vivo genetic toolkit for studying expression and functions of *Drosophila melanogaster* microRNAs. *RNA Biology*. 2016 (just-accepted):00-00
- [194] Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNADB: A reference database for long noncoding RNAs. *Nucleic Acids Research*. 2011;**39**(Suppl 1):D146-D151
- [195] Bhartiya D, Pal K, Ghosh S, Kapoor S, Jalali S, Panwar B, Jain S, Sati S, Sengupta S, Sachidanandan C et al. lncRNome: A comprehensive knowledgebase of human long non-coding RNAs. *Database*. 2013;bat034
- [196] Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*. 2014;**42**(D1):D92-D97
- [197] Consortium TR. RNAcentral: A comprehensive database of non-coding RNA sequences. *Nucleic Acids Research*. 2017;**45**(D1):D128-D134
- [198] Weikard R, Demasius W, Kuehn C. Mining long noncoding RNA in livestock. *Animal Genetics*. 2016
- [199] Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Molecular Cell*. 2011;**43**(6):904-914
- [200] Quagliata L, Matter MS, Piscuoglio S, Arabi L, Ruiz C, Procino A, Kovac M, Moretti F, Makowska Z, Boldanova T. Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. *Hepatology*. 2014;**59**(3):911-923
- [201] Liao Q, Xiao H, Bu D, Xie C, Miao R, Luo H, Zhao G, Yu K, Zhao H, Skogerbo G et al. ncFANs: A web server for functional annotation of long non-coding RNAs. *Nucleic Acids Research*. 2011;**39**(Suppl 2):W118-W124
- [202] Laible G, Wei J, Wagner S. Improving livestock for agriculture - technological progress from random transgenesis to precision genome editing heralds a new era. *Biotechnology Journal*. 2015;**10**(1):109-120
- [203] Anamika K, Verma S, Jere A, Desai A. Transcriptomic Profiling Using Next Generation Sequencing—Advances, Advantages, and Challenges. In: Kulski JK, editor. *Next Generation Sequencing - Advances, Applications and Challenges*. 2016. Rijeka: InTech. Ch. 04



- [204] Veneziano D, Nigita G, Ferro A. Computational approaches for the analysis of ncRNA through deep sequencing techniques. *Frontiers in Bioengineering and Biotechnology*. 2015;**3**:77
- [205] Proudfoot C, Carlson DF, Huddart R, Long CR, Pryor JH, King TJ, Lillico SG, Mileham AJ, McLaren DG, Whitelaw CB et al. Genome edited sheep and cattle. *Transgenic Research*. 2015;**24**(1):147-153
- [206] Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: Progress, implications and challenges. *Human Molecular Genetics*. 2014;**23**(R1):R40-R46
- [207] Yu L, Batara J, Lu B. Application of Genome Editing Technology to MicroRNA Research in Mammals. In: *Modern Tools for Genetic Engineering*, Michael Kormann (Ed.), InTech, Ch. 7, DOI: 10.5772/64330
- [208] Cox DBT, Platt RJ, Zhang F. Therapeutic genome editing: Prospects and challenges. *Nature Medicine*. 2015;**21**(2):121-131
- [209] Kevan MA Gartland MD, Tommaso B, Mariapia VM and Jill SG. Advances in biotechnology: Genomics and genome editing. *The EuroBiotech Journal*. 2017;**1**(1):3-10
- [210] Shen S, Loh TJ, Shen H, Zheng X, Shen H. CRISPR as a strong gene editing tool. *BMB Reports*. 2017;**50**(1):20-24
- [211] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;**157**(6):1262-1278
- [212] Zhuo C, Hou W, Hu L, Lin C, Chen C, Lin X. Genomic editing of non-coding RNA genes with CRISPR/Cas9 ushers in a potential novel approach to study and treat schizophrenia. *Frontiers in Molecular Neuroscience*. 2017;**10**:28
- [213] West J, Gill WW. Genome Editing in Large Animals. *Journal of Equine Veterinary Science*. 2016;**41**:1-6
- [214] Petersen B, Niemann H. Molecular scissors and their application in genetically modified farm animals. *Transgenic Research*. 2015;**24**(3):381-396
- [215] Tan WS, Carlson DF, Walton MW, Fahrenkrug SC, Hackett PB. Precision editing of large animal genomes. *Advances in Genetics*. 2012;**80**:37-97
- [216] Zhang JH, Adikaram P, Pandey M, Genis A, Simonds WF. Optimization of genome editing through CRISPR-Cas9 engineering. *Bioengineered*. 2016;**7**(3):166-174
- [217] Wang X, Zhou J, Cao C, Huang J, Hai T, Wang Y, Zheng Q, Zhang H, Qin G, Miao X et al. Efficient CRISPR/Cas9-mediated biallelic gene disruption and site-specific knockin after rapid selection of highly active sgRNAs in pigs. *Scientific Reports*. 2015;**5**:13348
- [218] Whitworth KM, Rowland RRR, Ewen CL, Tribble BR, Kerrigan MA, Cino-Ozuna AG, Samuel MS, Lightner JE, McLaren DG, Mileham AJ et al. Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. *Nature Biotechnology*. 2016;**34**(1):20-22



- [219] Butler JR, Ladowski JM, Martens GR, Tector M, Tector AJ. Recent advances in genome editing and creation of genetically modified pigs. *International Journal of Surgery* (London, England). 2015;**23**(Pt B):217-222
- [220] Lillico SG, Proudfoot C, Carlson DF, Stverakova D, Neil C, Blain C. Live pigs produced from genome edited zygotes. *Scientific Report*. 2013;**3**:2847
- [221] Wang K, Ouyang H, Xie Z, Yao C, Guo N, Li M, Jiao H, Pang D. Efficient generation of myostatin mutations in pigs using the CRISPR/Cas9 system. *Scientific Report*. 2015;**5**:16623
- [222] Choi W, Yum S, Lee S, Lee W, Lee J, Kim S, Koo O, Lee B, Jang G. Disruption of exogenous eGFP gene using RNA-guided endonuclease in bovine transgenic somatic cells. *Zygote* (Cambridge, England). 2015;**23**(6):916-923
- [223] Carlson DF, Lancto CA, Zang B, Kim E-S, Walton M, Oldeschulte D, Seabury C, Sonstegard TS, Fahrenkrug SC. Production of hornless dairy cattle from genome-edited cell lines. *Nature Biotechnology*. 2016;**34**(5):479-481
- [224] Crispo M, Mulet AP, Tesson L, Barrera N, Cuadro F, dos Santos-Neto PC, Nguyen TH, Creneguy A, Brusselle L, Anegon I et al. Efficient Generation of Myostatin Knock-Out Sheep Using CRISPR/Cas9 Technology and Microinjection into Zygotes. *PLoS One*. 2015;**10**(8):e0136690
- [225] Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*. 2016;**34**(9):933-941
- [226] Pulido-Quetglas C, Aparicio-Prat E, Arnan C, Polidori T, Hermoso T, Palumbo E, Ponomarenko J, Guigo R, Johnson R. Scalable design of paired CRISPR guide RNAs for genomic deletion. *PLOS Computational Biology*. 2017;**13**(3):e1005341
- [227] Wu B, Luo L, Gao XJ. Cas9-triggered chain ablation of cas9 as a gene drive brake. *Nature Biotechnology*. 2016;**34**(2):137-138
- [228] Gonen S, Jenko J, Gorjanc G, Mileham AJ, Whitelaw CBA, Hickey JM. Potential of gene drives with genome editing to increase genetic gain in livestock breeding programs. *Genetics Selection Evolution*. 2017;**49**(1):3
- [229] Aparicio-Prat E, Arnan C, Sala I, Bosch N, Guigó R, Johnson R. DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. *BMC Genomics*. 2015;**16**(1):846
- [230] Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*. 2015;**33**(5):510-517